Integrated Circuits and Systems

Christian C. Enz
Andreas Kaiser *Editors*

MEMS-based Circuits and Systems for Wireless Communication

Springer

Integrated Circuits and Systems

Christian C. Enz • Andreas Kaiser
Editors

MEMS-based Circuits and Systems for Wireless Communication

Springer

Editors
Christian C. Enz
CSEM SA
CH-2002 Neuchâtel
Switzerland
christian.enz@csem.ch

Andreas Kaiser
IEMN, Département ISEN
59046 Lille
France
andreas.kaiser@isen.fr

ISSN 1558-9412
ISBN 978-1-4419-8797-6 ISBN 978-1-4419-8798-3 (eBook)
DOI 10.1007/978-1-4419-8798-3
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012943350

Preface

Over many years, RF-MEMS have been a hot topic in research at the technology and device level. In particular, various kinds of mechanical Si-MEMS resonators and piezoelectric BAW (bulk acoustic wave) resonators have been developed. The BAW technology has made its way to commercial products for passive RF filters, in particular for duplexers in RF transceiver front ends for cellular communications. Beyond their use in filters, micromachined resonators can also be used in conjunction with active devices in innovative circuits and architectures. Possible applications are active tunable RF front-end filters, frequency synthesizers for LO generation, or temperature-compensated MEMS resonators for frequency/time reference potentially replacing the long-time used quartz crystal. Furthermore, MEMS devices can advantageously be used in radios for further miniaturization and reduction of power consumption.

This book presents a broad overview of this technology going from the MEMS devices, mainly BAW and Si-MEMS resonators, to basic circuits such as oscillators and finally complete systems such as ultralow-power MEMS-based radios. The work is targeted at circuit and system designers. The fabrication process of the MEMS devices is only covered at a minimal level. The discussion of MEMS devices focuses on their properties and modeling, so they can be efficiently used in circuits. Circuit design specific to MEMS devices is discussed in depth. Traditional circuits cannot be used with high-Q resonators, and special techniques for oscillator and filter design are required. Finally, several examples of system architectures built around MEMS devices are described. It is particularly shown how these architectures can exploit the potential of the MEMS devices to reduce size and power consumption for applications such as wireless sensors where these parameters are critical.

The book is organized in three parts. The first part considers devices, models, and passive circuits. Dubois et al. briefly introduce in the first chapter the BAW (bulk acoustic wave) technology and describe in detail the modeling of BAW resonators. Model complexity depends on the range of phenomena that need to be considered, and equivalent circuit level models for BAW resonators are developed. The second chapter by *Piazza* focuses on a particular class of resonators using contour-mode resonance. This allows adjustment of the resonance frequency at

mask level as opposed to the FBAR or SMR resonators where the resonance frequency is determined at the technology level. Several examples of passive circuits designed with this approach are given. The following two chapters introduce more prospective aspects. *Ionescu* gives a large overview of the state-of-the-art and the ongoing developments of nanoelectromechanical systems (NEMS) relevant to communication circuits. Numerous examples of passive and active devices such as nanowires, nanotubes, NEMS switches, mixers, and active resonators are shown as well as their conceptual use in radios. Starting from the physical properties of acoustic devices, *Dubus* describes how these properties could be used in various ways to increase functionality of acoustic devices. Resonators could be made tunable at the device level, and applications such as frequency-based multiplexing and demultiplexing could be implemented with phononic crystals.

The second part of the book is dedicated to circuits using BAW resonators. *Vittoz* gives in Chap. 5 a detailed treatment of high-Q crystal oscillator design and describes the different known topologies from a theoretical point of view. *Tournier* describes in Chap. 6 several practical implementations of oscillators in BiCMOS technology with above-IC FBAR resonators. The following chapter by Ray et al. describes differential quadrature CMOS/BAW oscillators for LO generation in very low power applications making use of control loops for temperature compensation and phase error correction. In the last chapter of Part II, *Razafimandimby et al.* present tunable BAW filters employing active Q-enhanced inductors and negative capacitance circuits. A semidigital control loop adapted to the BAW filter context allows precise frequency tuning.

The third part of the book presents various systems using RF-MEMS as key components. Otis et al. present various possibilities of using BAW resonators for impedance matching, tuned amplifiers, and image reject transformers. These circuits are used in a complete superregenerative BAW-based receiver for asynchronous communications as well as a BAW-based ultralow-power wake-up receiver with uncertain IF. In the following chapter, *Ruffieux* describes another original radio architecture using Si and BAW resonators for frequency reference, LO generation, and filtering combined with an all-digital phase locked loop. Ito et al. introduce the use of BAW oscillators as digitally controlled frequency reference calibrating itself, thanks to information transmitted on the radio network. Finally, a complete wireless sensor node for tire pressure monitoring in automotive applications is described by Dielacher et al. in Chap. 12. The system is built around a MEMS sensor and a BAW-based CMOS RF transmitter for ultralow-power consumption and employs advanced packaging technologies.

As can be seen from the contributions presented in this book, RF-MEMS and particularly BAW resonators are about to become key components in RF transmitters. This trend will certainly continue with the growing need for ultralow-power radios in areas including sensor networks, body area networks, and automation of homes and offices.

Neuchâtel, Switzerland — Christian Enz
Lille, France — Andreas Kaiser

Contents

Contributors

Andreia Cathelin STMicroeletronics, Crolles, France

Jérémie Chabloz CSEM, Centre Suisse d'Electronique et de Microtechnique, Neuchâtel, Switzerland

Matteo Contaldo CSEM, Centre Suisse d'Electronique et de Microtechnique, Neuchâtel, Switzerland

Markus Dielacher Infineon Technologies, Graz, Austria

Marc-Alexandre Dubois Swiss Center for Electronics and Microtechnology (CSEM S.A.), Neuchâtel, Switzerland

Bertrand Dubus Institut d'Electronique de Microélectronique et de Nanotechnologie, Département ISEN, Lille, France

Christian C. Enz CSEM, Centre Suisse d'Electronique et de Microtechnique, Neuchâtel, Switzerland

Martin Flatscher Infineon Technologies, Graz, Austria

Thomas Herndl Infineon Technologies, Graz, Austria

Adrian Ionescu Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Hiroyuki Ito Tokyo Institute of Technology, Yokohama, Japan

Andreas Kaiser Institut d'Electronique, de Microélectronique et de Nanotechnologie, Département ISEN, Lille, France

Hasnain Lakdawala Intel Corporation, Hillsboro, OR, U.S.A.

Thomas Lentsch Infineon Technologies, Graz, Austria

Rainer Matischek Infineon Technologies, Graz, Austria

Claude Muller Swiss Center for Electronics and Microtechnology (CSEM S.A.), Neuchâtel, Switzerland

Brian P. Otis University of Washington, Seattle, WA, U.S.A.

Gianluca Piazza University of Pennsylvania, Philadelphia, PA, U.S.A.

Nathan Pletcher Qualcomm Incorporated, San Diego, CA, U.S.A.

Josef Prainsack Infineon Technologies, Graz, Austria

Jan Rabaey University of California, Berkeley, CA, U.S.A.

Shailesh S. Rai University of Washington, Seattle, WA, U.S.A.

Ashoke Ravi Intel Corporation, Hillsboro, OR, U.S.A.

Stéphanne Razafimandimby STMicroeletronics, Crolles, France

David Ruffieux Swiss Center for Electronics and Microtechnology (CSEM S.A.), Neuchâtel, Switzerland

Cyrille Tilhac STMicroeletronics, Crolles, France

Éric Tournier LAAS/CNRS, Université de Toulouse, Toulouse, France

Éric A. Vittoz Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Werner Weber Infineon Technologies, Munich, Germany

Acronyms

AlN	Aluminum nitride
A0, A1, A2	Antisymmetrical lamb waves
BAW	Bulk acoustic wave
BST	Barium strontium titanate
BTO	Barium titanate
BW	Bandwidth
DCS	Digital cellular system
FBAR	Film bulk acoustic resonator
GSM	Global system for mobile communications
IDT	InterDigitated transducer
IF	Intermediate frequency
IL	Insertion loss
KLN	Potassium lithium niobate
KNO	Potassium niobate
LNO	Lithium niobate
LTO	Lithium tantalate
MEMS	Micro-electromechanical system
PC	Phononic crystal
PCS	Personal communications service
PMN	Lead magnesium niobate
PT	Lead titanate
PZT	Lead zirconate titanate
RF	Radio frequency
RL	Rejection level
SAW	Surface acoustic wave
SH	Shear horizontal
SMR	Solidly mounted resonator
STO	Strontium titanate
S0, S1, S2	Symmetrical lamb waves
TE	Thickness extensional

TS	Thickness shear
TS2	First harmonic of thickness shear
UHF	Ultra high frequency
W-CDMA	Wideband code division multiple access evaluation
ZnO	Zinc oxide

Part I
NEMS/MEMS Devices

Chapter 1
Thin-Film Bulk Acoustic Wave Resonators

Marc-Alexandre Dubois and Claude Muller

Abstract Miniature bulk acoustic wave (BAW) resonators are components that exhibit very interesting properties for communication systems, as confirmed by their extensive use nowadays in front-end filters for mobile phones. This chapter reviews the technology enabling the fabrication of these devices and the different models used to describe their electrical performances. Finally, a simple empirical model, mainly based on geometrical parameters, is proposed. It does not require massive computing power, but it can nevertheless predict very accurately the main characteristics of the thin-film BAW resonators.

1.1 Introduction

Many electronic systems rely on their ability to select or generate signals with a very precise frequency. Hence, they require filters for sorting the right signals among others and oscillators for providing a stable reference frequency. The common feature of these blocks is their use of resonators, of which performance is extremely important, especially in the case of low-noise or low-power designs. Indeed, the quality factor Q of the resonator determines the insertion loss of the filter, or the phase noise of the oscillator.

Among the different techniques available for making a resonator, exploiting the propagation of acoustic waves in a solid medium is the best way to create a compact device. This is due to the much lower phase velocity of the acoustic wave—approximately five orders of magnitude—compared to the velocity of an electromagnetic wave. At a given frequency, the size of the resonating element in the

M.-A. Dubois (✉) • C. Muller
Centre Suisse d'Electronique et de Microtechnique (CSEM), Neuchâtel, Switzerland
e-mail: marc-alexandre.dubois@csem.ch; claude.muller@csem.ch

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_1,

acoustic resonator can hence be made much smaller than, for example, the minimum length of coaxial line or coplanar wave guide required by an EM resonator.

Even though there is a large variety of acoustic waves, each of these featuring its own characteristics and propagation mode (see Chap. 4), resonators are usually referred to only according to two coarse categories, BAW and SAW: when the acoustic wave is propagating in the bulk of the material composing the device, while occupying all or most of its volume, it is called a bulk acoustic wave (BAW), as opposed to a wave trapped and traveling at the interface between the solid and the air, which is a surface acoustic wave (SAW). The resonators described in this chapter are from the BAW category.

In order for a BAW resonator to work properly, the acoustic wave propagating in the solid has to be confined within the volume of the resonator itself. In other words, the acoustic energy has to be trapped locally so as not to leak out of the device. This is done by introducing discontinuities in the path of the acoustic wave so that the latter is reflected. The most efficient discontinuity is the simple air–solid interface, but other ways exist, such as Bragg reflecting stacks. Another requirement for a good resonator is that the medium of propagation itself should not dissipate too much energy, for example, through viscoelastic losses. The material should hence be chosen carefully.

So a BAW resonator can be seen as a volume of material in which an acoustic wave is bouncing back and forth between reflecting interfaces. But how is the acoustic wave generated in the first place?

One elegant way to perform this is to resort to piezoelectric materials. Piezoelectricity is a phenomenon exhibited by some materials, most of which being crystalline, which couples their mechanical and electrical properties. The capability of these solids to develop an electric polarization when they are strained through mechanical stress is called the direct piezoelectric effect. It is due to the fact that their crystal structure lacks a center of symmetry, so that an applied stress gives rise to an asymmetrical ionic displacement, and hence to a net change in dipole moment. The same materials display also the converse piezoelectric effect: when an electric field is applied to them, they change their dimensions.

A simple and practical transducer for generating acoustic waves can thus be made of a slab of piezoelectric crystal coated with metal electrodes. The application of a sinusoidal voltage to the electrodes will result in a periodic deformation of the crystal. If the transducer is not in contact with another solid, to which this acoustic wave can be transmitted, i.e., its vibrating part is surrounded only by a low acoustic impedance medium such as air or vacuum, it is also a resonator, of which resonance frequency is determined by its dimensions.

An important parameter regarding piezoelectricity is the amount of electrical energy that is converted to mechanical energy, and vice versa. The level of this electro-acoustic conversion is described by the piezoelectric coupling coefficient K^2, which is a material property. This parameter is crucial for the designer since it determines the maximum bandwidth achievable for a filter composed of BAW resonators, or the frequency range in which a BAW-based oscillator can be tuned.

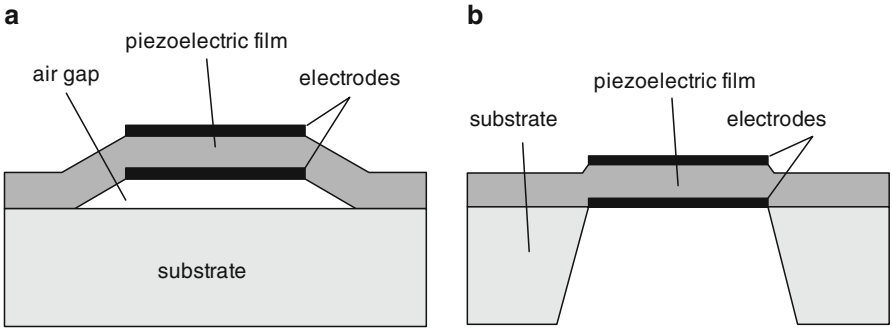

Fig. 1.1 Cross sections of FBARs realized by (**a**) surface micromachining or (**b**) bulk micromachining

Among the large family of piezoelectric materials, quartz crystals have always been preferred for making BAW resonators. Apart from being piezoelectric, they are able to sustain acoustic waves with very limited damping, and some crystal cuts exhibit an extremely small sensitivity to temperature variations. The latter property is extremely valuable if the resonator is to be used as a frequency reference. Other examples of piezoelectric materials of interest for the resonators industry include single crystals of lithium tantalate or lithium niobate—mainly for SAW applications—and aluminum nitride (AlN) or zinc oxide (ZnO) in thin-film form.

1.1.1 Thin-Film Bulk Acoustic Wave Resonators

A thin-film BAW resonator is a device composed mainly from a piezoelectric thin film surrounded by two metal electrodes that generates an acoustic wave propagating according to a thickness mode. The way the wave is trapped in the resonator is the main differentiator between the two types of devices that have reached volume manufacturing today.

The first one uses air–solid interfaces both over and underneath the resonating film. The resonator—also known as *film bulk acoustic resonator* or FBAR—is hence a membrane suspended in air by its edges. It can be manufactured over a sacrificial layer, or by etching part of the substrate underneath the resonator (Fig. 1.1).

The second configuration is the *solidly mounted resonator* or SMR: it is a more robust structure where the top interface is also of the air–solid type, but which uses an acoustic reflector as bottom interface. An efficient isolation is performed owing to the transformation of the acoustic impedance of the substrate over which the resonator is built, to a very low value, through a set of quarter-wavelength sections of materials having different elastic properties (Fig. 1.2). The more different these properties are from each other, the smaller is the number of layers required in the

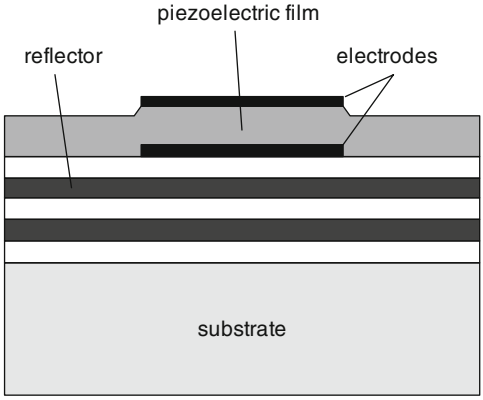

Fig. 1.2 Cross section of a solidly mounted resonator

reflector. For example, the use of AlN and SiO_2 requires at least nine alternating layers, whereas replacing AlN by tungsten allows this number to be reduced down to five.

1.1.2 Background

Thin-film BAW resonators are an answer to the ever increasing operating frequency of modern communication systems. As the resonance frequency of a BAW resonator working at its fundamental mode is determined by the size of the acoustic confinement structure—which is half a wavelength long—the regular quartz technology cannot be applied in the GHz domain. Even though fabrication methods have been developed for manufacturing high-frequency inverted mesa resonators, by locally thinning down quartz plates, this technology finds its limit in the 250-MHz range.

It was recognized early on that instead of thinning down a piezoelectric plate to unpractical values, depositing a thin layer of piezoelectric material might be a more suitable method to reach higher operation frequencies [1]. However, more than a decade of technology development was still ahead before this new concept could be successfully demonstrated. Progress was required first in the growth of good quality piezoelectric films on metal electrodes, which spurted the study of AlN and ZnO sputtering methods, but also in the field of patterning and the process technologies that are typical from the integrated circuit industry—including photolithography, magnetron sputtering of metal films, wet and dry etching—that were still in the development phase.

The initial developments focused on bulk micromachined, membrane-type resonators, called at the time *composite resonators* because they still required a rather thick layer of silicon or silicon oxide underneath the piezoelectric film, for strengthening the membrane [2–4]. As a consequence, these devices operated in the few hundreds of MHz range, while their coupling coefficient was somewhat limited by the presence of this additional material.

Then, the first FBARs without the Si supporting layer in the membrane, hence featuring a much larger coupling coefficient [5], and the use of surface micromachining [6] appeared as a natural evolution. SMRs were however demonstrated only a decade later [7]. From that time, owing to the craving of the mobile phone industry for small duplexers meeting the tough specifications of the new standards around 2 GHz, the momentum in research and development of the thin-film BAW technology was tremendously increased. Aside from the design of efficient resonators and filters, much effort was spent to bring the fabrication processes to volume manufacturing standards. FBAR filters arrived on the mobile phone market just after the turn of the century [8], soon followed by their SMR cousins [9].

1.2 Technology

The fabrication of thin-film BAW resonators and filters is quite complicated, mainly due to the many different layers composing the devices. Since most of these layers need some type of patterning, the number of required process steps is high. Moreover, unlike in many other devices, most layers in the resonator have several functions, for example, both acoustical and electrical: a metal electrode does not only need to bring current to the resonator, it also takes part in enhancing the effective coupling coefficient, and it contributes to the trapping of the acoustic wave. Consequently, several material parameters—such as resistivity, stress, or surface roughness—need to be optimized simultaneously for each film.

The very large sensitivity of the resonance frequency to any thickness variation is but another difficulty specific to the BAW manufacturing process. It could be overcome only through dedicated developments of sputtering systems by some equipment manufacturers [10]. Thickness uniformity across a wafer has been narrowed down by nearly a factor of ten, compared to what was used in the microelectronics industry. And still trimming the resonators through ion beam etching cannot be avoided to maintain production yields at an economically viable level. All this places the FBARs and SMRs among the most demanding components in terms of process control.

1.2.1 Aluminum Nitride

The heart of the BAW resonator is the piezoelectric thin film. After the first years of development, during which ZnO was very much used, aluminum nitride (AlN) emerged as the most suitable technology for BAW resonators because it is an excellent compromise between performance and manufacturability. Its coupling coefficient is not as high as that of ZnO or PZT, but it is chemically very stable, with a bonding energy of 11.5 eV, and it benefits from an excellent

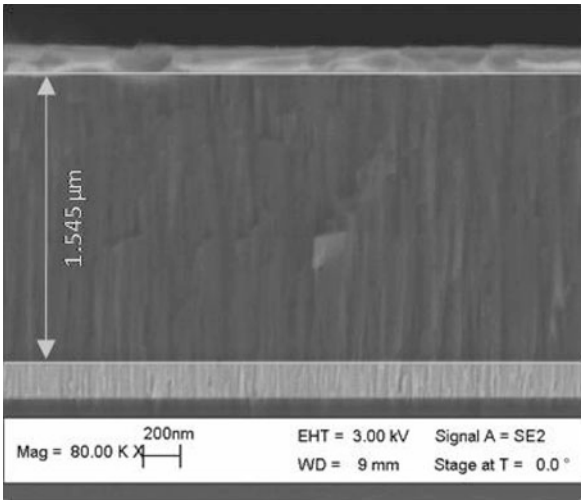

Fig. 1.3 SEM cross section of AlN film between metal electrodes

thermal conductivity and a low temperature coefficient. These properties enable the fabrication of resonators featuring coupling factors of 6–7%, good resistance to corrosion, excellent power handling capability, and limited drift with temperature. Another advantage of AlN is the low process temperature and the fact that it does not contain any contaminating elements harmful for semiconductor devices, unlike most other piezoelectric materials. This is essential in the case of monolithic integration of BAW resonators with microelectronic integrated circuits [11].

Reactive sputtering from a pure Al target in a plasma containing nitrogen is the most suitable method to obtain crystalline AlN films with sufficient quality for BAW applications. Both AC and pulsed DC power supplies can be used to sustain the plasma. Figure 1.3 shows the cross section of such an AlN film grown in pulsed DC mode. The microstructure is typical, with very densely packed columnar grains.

The parameters of the AlN deposition process have to be optimized in order to ensure that a vast majority of grains are oriented along the c-axis since the spontaneous polarization of AlN, and hence the maximum piezoelectric effect, is parallel to that direction. In addition to the process parameters, the bottom electrode, which acts as a seeding layer for AlN, is equally important regarding the piezoelectric properties of the film. It is mandatory that the surface of this electrode be extremely smooth, and also free of oxygen. An ion milling or sputter-etching step of the surface prior to the AlN deposition is the usual way to reach good nucleation conditions. BAW resonators with high coupling have been demonstrated using electrode of platinum, aluminum, molybdenum, tungsten, or even ruthenium.

1.2.2 Process Flow for SMRs

There are many different possible process flows for manufacturing thin-film BAW resonators. Each company active in this field has come with its own, which is often

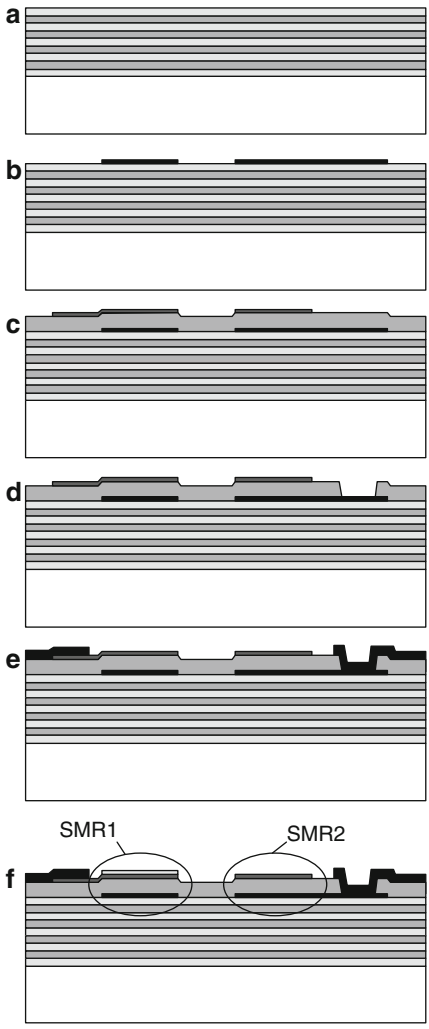

Fig. 1.4 Example of simple process flow for solidly mounted resonators. The resonator SMR1 has a lower resonance frequency than SMR2, due to additional loading

the result of many years of development, and hence is kept secret for obvious reasons. Figure 1.4 is a very simple example of process flow for SMRs. Starting from a bare silicon wafer, a set of alternating quarter-wavelength layers of SiO_2 and AlN are deposited, to serve as acoustic reflector (a). Then, a Pt bottom electrode is sputter-deposited and patterned by dry etching (b). This is followed by the deposition of the piezoelectric AlN thin film and its top Al electrode, which is also patterned by dry etching (c). Next, via holes are dry etched into AlN, to get access to the bottom electrode (d), and an interconnection metallic layer (Al) is first deposited and then patterned (e). This Al interconnect is used both as pad metal for bonding or probing the devices and for connecting different resonators together into a filter architecture. Finally, a SiO_2 loading layer is sputter-deposited and patterned, for

lowering the resonance frequency of some resonators (f). This last step is required to build lattice or ladder filters, which need resonators with two slightly different resonance frequencies.

This process flow can be kept simple by the fact that the materials used for the acoustic reflector are dielectric. In the case where metal is used, such as high-impedance tungsten in the well-known SiO_2/W combination, a much more complicated fabrication scheme has to be devised. Indeed, the reflector needs to be patterned under each resonator to prevent any electrical cross talk from one resonator to the next through the layers in the reflector.

1.3 Modeling BAW Resonators

1.3.1 Spurious Modes

Electronic applications normally require the largest possible coupling coefficient and Q factors. Besides these two fundamental requirements, it is also very desirable to have a very smooth response. For example, only small ripples are accepted in the passband of a filter. Depending on the design and technology, BAW devices can show a very smooth response or lots of ripples. The origin of the ripples is the presence of spurious resonance modes which are excited simultaneously with the main mode of the BAW resonators. These weakly coupled spurious modes superimpose the main mode and create ripples.

The study of the spurious modes implies a deep understanding of the wave mechanics [12–14]. This is out of the scope of this chapter (please refer to Chap. 4). Only a few basic concepts will be covered here. Depending on their properties, mechanical waves (sound waves) are classified in different categories. Bulk waves are classified in ten groups, each group corresponding to a particular symmetry case of the particles motion. Three of them are fundamental groups: the dilatation group, the shear group, and the torsion group. The seven other groups are combinations, through coupling, of the three first groups, for example, the flexure group of thin plates, or the contour mode group used in quartz resonators. Besides bulk waves, there are also surface waves that are utilized in SAW devices for example. Well-known surface wave types are Love waves and Rayleigh waves. Finally, there are also plate waves like the Lamb waves. They only propagate in plates, or in other terms in a wave guide. These waves are of particular importance for BAW resonators since any thin film used in the BAW stack can play the role of a wave guide.

In a resonator, the traveling waves are reflected at the boundaries of the resonator, so as to create a standing wave called resonance mode. Depending on the geometry of the resonator and on the thickness and mechanical properties of the layers composing the resonator, particular modes are favored and others are completely killed. The aim with thin-film BAW resonators is to favor the fundamental dilatation mode while suppressing all the other modes. In practice, this is a very difficult

task, due to the huge number of modes that can potentially occur in a structure. However, a solution was proposed by a research group from Infineon/VTT/Nokia. It is based on the fact that, in BAW resonators, the most important spurious modes are standing Lamb waves that arise because of the boundary conditions on the wall of the resonator. The idea is to introduce an acoustic impedance matching layer at the edge of the resonator, so as to modify the boundary conditions and kill the Lamb waves. Experimental data show that the technique works well and spurious-free resonators are obtained. These resonators show very high Q with values in the 1,500–2,000 range. For more details, the interested reader can refer to [15] and [16].

1.3.2 One-Dimensional Mason Model

The most famous model used in the BAW field is the Mason model, which is a 1D model. As such, it cannot handle the spurious modes problem, which is intrinsically a 2- or 3D problem. However, it is a very useful model to design BAW devices. The Mason model allows calculating with a good precision the resonance frequency of a BAW resonator fabricated with a given stack of layers. It also gives an upper limit to the coupling coefficient and Q factor that can be achieved. Again, the real coupling and Q-factor that are achieved in a real device cannot be predicted by the 1D Mason model since it depends on the complete 3D geometry of the resonator. We will present in a later section a model that allows taking into account the 2D planar geometry of the resonator.

The Mason model relies on a rigorous treatment of the propagation of mechanical waves in a stack of infinitely large layers. Infinitely large layers are a way to get rid of lateral boundary conditions. In other terms, the problem is reduced to a 1D system for waves propagating perpendicular to the layers surfaces. For the non-piezoelectric layers (electrodes, SMR reflector, etc.), it is a simple problem of propagation of sinusoidal waves with continuity conditions at the interface between layers. Inside a layer, the velocity of the wave depends on the rigidity and the density of the layer. Continuity conditions on the displacement and stress at the interfaces between layers dictate the amplitude ratio between the transmitted wave and the reflected wave.

The treatment of the piezoelectric layer is more complicated because of the electromechanical coupling. The continuity conditions on the displacement and stress at the interfaces of the piezoelectric layer do not depend solely on mechanical terms anymore. They also depend on the electrical potential at these two interfaces. The system is best described in a matrix form, where a submatrix addresses the purely mechanical part, one term addresses the purely electrical part, and the remaining terms account for the electromechanical coupling [14].

The solution for a given device is obtained by cascading the various non-piezoelectric and piezoelectric layers as they appear in the device, and to solve the corresponding set of equations for the global system. The Mason model being analytical and relying on no approximations, it can be applied from the most simple

structure, such as a freestanding quartz crystal, to much more complicated systems, like SMR BAW resonators or ultrasonic transducers for medical imaging.

As an illustration, and since it introduces in an easy way a few basic concepts linked to piezoelectric resonators, the remaining of this section is dedicated to the Mason model applied to a freestanding resonator with electrodes so thin they can be neglected. The resonator is then simply a piezoelectric plate surrounded by two media having the same or different acoustic impedances. This case is presented in many textbooks [13,14]. The reader will refer to them for the complete mathematical developments. Only the major results are given hereafter.

From the Mason model, it can be shown that the electrical impedance of the resonator is given by:

$$Z_e = \frac{1}{j\,\omega C_0}\left(1 + \frac{K^2}{\phi} Z_p \frac{2Z_p\,(1-\cos\phi) - j\,(Z_1+Z_2)\sin\phi}{-\left(Z_p^2 + Z_1 Z_2\right)\sin\phi + j\,Z_p\,(Z_1+Z_2)\cos\phi}\right), \tag{1.1}$$

where Z_p is the elastic impedance of the piezoelectric material, ω is the angular frequency, C_0 is the dielectric capacitance, K^2 is the electromechanical coupling coefficient, Z_1 and Z_2 are the elastic impedance of the materials on each side of the resonator, and $\phi = \omega d/v_p$, where d is the thickness of the piezoelectric plate and v_p is the wave velocity in the piezoelectric material.

The piezoelectric nature of the plate is expressed by the term proportional to K^2. Without it, (1.1) comes back to the electrical impedance of a simple dielectric layer.

A freestanding resonator is a resonator surrounded by layers having acoustic impedance equal to zero. In other terms, it stands in vacuum. Inserting $Z_1 = Z_2 = 0$ into (1.1), the electrical impedance becomes:

$$Z_e = \frac{1}{j\,\omega C_0}\left(1 - K^2 \frac{\tan(\phi/2)}{\phi/2}\right). \tag{1.2}$$

As shown in Fig. 1.5, the impedance curve of the resonator corresponds to the impedance of a capacitance C_0 on which resonances are superimposed. In the case under study, there is no loss and impedance is purely imaginary. At the antiresonance frequencies, it is infinite. The anti-resonance frequencies f_a are given by (1.2) when $\phi = (2n+1)\pi/2$. With $\phi = \omega d/v_p$, it follows that the first antiresonance frequency is given by

$$f_a = \frac{v_p}{2d}. \tag{1.3}$$

It is simply the frequency for which the thickness of the plate corresponds to a half wavelength.

The resonance frequencies are given by (1.2) when the electrical impedance is zero. This condition is met when

$$K^2 \frac{\tan(\phi/2)}{\phi/2} = 1. \tag{1.4}$$

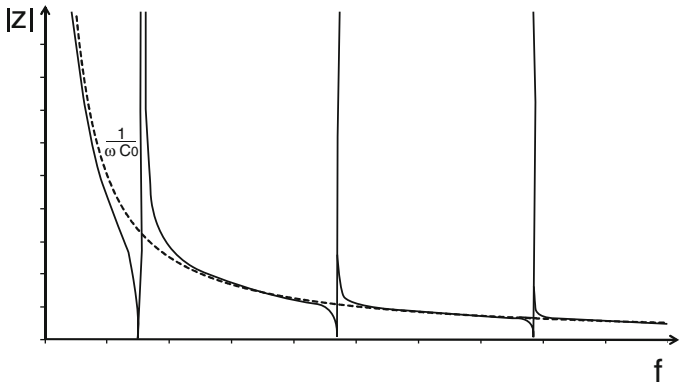

Fig. 1.5 Impedance curve of a freestanding resonator whose impedance is given by (1.2). The dashed line represents a simple dielectric capacitance

Using $\phi = \omega d/v_p$ and $v_p = f_a \cdot 2d$, it can be rewritten as

$$K^2 \tan\left(\frac{\pi}{2}\frac{f_r}{f_a}\right) = \frac{\pi}{2}\frac{f_r}{f_a}. \tag{1.5}$$

This is a transcendental equation that cannot be solved analytically. If necessary, numerical methods like Newton's method can be applied. However, the interest of (1.5) is not here. Equation (1.5) can be rewritten as

$$K^2 = \frac{\pi}{2}\frac{f_r}{f_a}\cot\left(\frac{\pi}{2}\frac{f_r}{f_a}\right). \tag{1.6}$$

In practice, (1.6) is used to determine the real coupling coefficient of a resonator. It is simply obtained from the resonance and antiresonance frequencies extracted from the impedance curve of the resonator. It is noteworthy that the larger the coupling coefficient is, the wider the separation between f_r and f_a.

This section about the Mason model is concluded with an approximation that will be used in the next section about the electrical equivalent circuit of a piezoelectric resonator. This approximation is based on the development of the function $tan(x)/x$ in a series of fractions

$$\frac{\tan x}{x} = \sum_{n=0}^{\infty} \frac{2}{[(2n+1)\pi/2]^2 - x^2}. \tag{1.7}$$

Keeping only the first term and introducing this into (1.2), one gets

$$Z_e \approx \frac{1}{j\omega C_0}\left(1 + K^2 \frac{2}{\frac{\pi^2}{4} - \frac{\phi^2}{4}}\right). \tag{1.8}$$

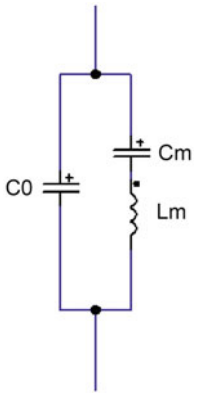

Fig. 1.6 Simple LC resonant circuit

This impedance is equivalent to the following admittance:

$$Y_e \approx \frac{j\,\omega\,C_0}{1-\frac{8\,K^2/\pi^2}{1-\omega^2/\omega_a^2}}, \quad \text{where} \quad \omega_a = \frac{\pi v_p}{d}, \tag{1.9}$$

or

$$Y_e \approx j\,\omega\,C_0\frac{\omega_a^2-\omega^2}{\omega_r^2-\omega^2} \quad \text{where} \quad \omega_r^2 = \omega_a^2\left(1-\frac{8K^2}{\pi^2}\right). \tag{1.10}$$

By taking only the first term in the series (1.7), the approximation is equivalent to considering only the main resonance and neglecting higher-order resonances. The approximation is thus only valid in the vicinity of the main resonance.

1.3.3 Electrical Equivalent Circuit (1D)

The circuit of Fig. 1.6 is composed of a static capacitance C_0 and of the motional capacitance and inductance C_m and L_m. Its admittance is given by

$$Y = j\,\omega\,C_0\left(1+\frac{C_m/C_0}{1-L_m C_m\,\omega^2}\right). \tag{1.11}$$

This circuit corresponds to a resonant circuit with a series resonance ω_s and a parallel resonance ω_p given respectively by

$$\omega_s^2 = \frac{1}{L_m C_m}, \tag{1.12}$$

$$\omega_p^2 = \omega_s^2\left(1+\frac{C_m}{C_0}\right). \tag{1.13}$$

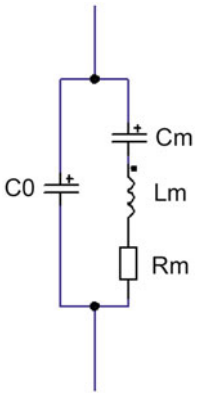

Fig. 1.7 Butterworth–Van Dyke (BVD) equivalent circuit

Using these two relations, one can easily transform (1.11) for the admittance into

$$Y = j\omega C_0 \frac{\omega_p^2 - \omega^2}{\omega_s^2 - \omega^2}. \tag{1.14}$$

The correspondence with (1.10) is evident. Thus, the equivalent circuit shown in Fig. 1.6 can be used to model the main resonance of a *low-frequency*, *freestanding*, *lossless* resonator. In the lossless case, the following equalities hold:

$$\omega_s = \omega_r \quad \text{and} \quad \omega_p = \omega_a.$$

Finally, the coupling coefficient can be calculated from the elements of the equivalent circuit. By comparing (1.9) with (1.11), the following relation can be derived:

$$K^2 = \frac{\pi^2}{8} \frac{C_m}{C_0 + C_m}. \tag{1.15}$$

Electronic designers are used to choose the electrical components of their circuit freely. Equation (1.15) has to be understood the reverse way: C_m is defined by the coupling coefficient! It is imposed by the materials and geometry used in the resonator. It is hence clearer to write (1.15) as

$$C_m = \frac{8K^2/\pi^2}{1 - 8K^2/\pi^2} C_0. \tag{1.16}$$

Real resonators are of course not lossless. Losses arise in the form of ohmic losses, dielectric losses, and acoustic losses of various origins. To take these losses into account, resistive elements must be introduced in the equivalent circuit. Figure 1.7 shows the well-known Butterworth–Van Dyke (BVD) equivalent circuit where resistance R_m is placed in series with the motional capacitance and inductance C_m and L_m.

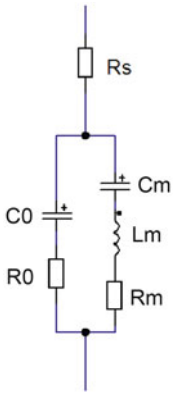

Fig. 1.8 Modified Butterworth–Van Dyke (MBVD) equivalent circuit with an additional series resistance R_s

This model has been extensively used in the quartz crystal field. With this equivalent circuit, the Q factors of the series resonance and of the parallel resonance are almost the same:

$$Q_s = \frac{L_m \omega_s}{R_m} \approx \frac{L_m \omega_p}{R_m} = Q_p.$$

Experimentally, Q_s and Q_p of thin-film BAW resonators are very often different. To take this into account, an additional resistance R_0 can be added in series with the static capacitance C_0. This model is called the modified Butterworth–Van Dyke (MBVD) equivalent circuit [17]. In Fig. 1.8, the additional series resistance R_s represents the resistance of the electrodes.

Now the Q factors at the resonance and at the antiresonance are different. They are given by

$$Q_s = \frac{L_m \omega_s}{R_m + R_s}, \tag{1.17}$$

$$Q_p = \frac{L_m \omega_p}{R_m + R_0}. \tag{1.18}$$

The main drawback of this model is that it contains three resistances while only two Q factors are extracted from measured curves. The choice of attributing the relative values among the three resistances is then quite arbitrary. A good way to solve this ambiguity is to use an EM simulator to get the R_s value for the resonator under consideration. Then R_m and R_0 are uniquely defined. Additionally, a third Q factor can be defined: the intrinsic series Q factor, Q_s^{intr}. It corresponds to the theoretical series Q factor, had the resonator perfect lossless connections and electrodes ($R_s = 0$). It gives an upper limit to the series Q factor that can be achieved. We will use it in a later section. It is given by

$$Q_s^{intr} = \frac{L_m \omega_s}{R_m}. \tag{1.19}$$

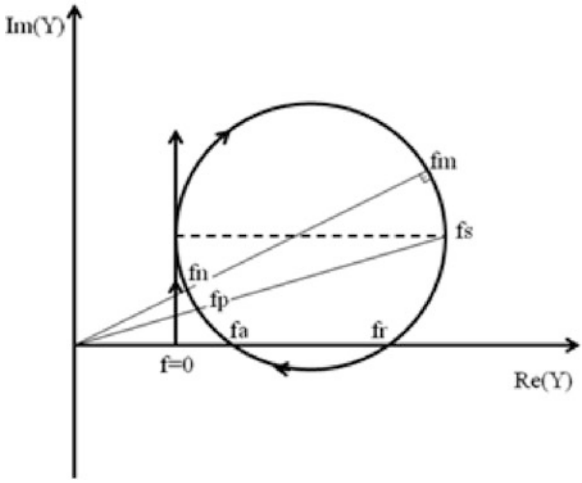

Fig. 1.9 Vector admittance diagram of a piezoelectric resonator

To conclude this section about electrical equivalent circuits, let's have a look at the resonance and antiresonance frequencies. In the lossless case, we found that $\omega_s = \omega_r$ and $\omega_p = \omega_a$. These equalities do not hold anymore when losses are accounted for. Both resonance and antiresonance frequencies are split into three different values:

- ω_s, ω_p: motional series, respectively parallel, frequency:

$$\omega_s = \frac{1}{\sqrt{L_m C_m}} \quad \text{and} \quad \omega_p = \omega_s \sqrt{1 - \frac{C_m}{C_0}}.$$

- ω_r, ω_a: resonance and antiresonance frequency, at which reactance is zero.
- ω_m, ω_n: frequency at which impedance is minimum, respectively maximum.

These six frequencies are shown on Fig. 1.9. The vector admittance curve of the resonator in the complex plane describes a vertical line with increasing frequency, except in the resonance region where it describes a circle.

1.3.4 2D/3D Models

As mentioned at the beginning of section on the Mason model, a 1D treatment can neither predict the real coupling and Q factor of a real resonator—since it depends on the complete 3D geometry of the device—nor handle the spurious modes (intrinsically a 2- or 3D problem). The requirements of today's telecommunication system are so severe that advanced optimization of the FBAR resonators is needed. Many groups developed 2D and 3D models to get a better understanding of

the physics governing these devices and so to improve their design. Most of these works are using FEM simulations and/or interferometric measurements of the displacement profile. The aim is to evaluate the coupling of other (spurious) modes such as Lamb- or Rayleigh-type lateral modes. A correct estimation of these couplings leads to much improved simulations, both in terms of ripples and performances (coupling and Q factors). The interested reader should refer to the original works, as, for example, [18–21].

In the remaining of this section, we will present a recently proposed model that allows an easy, intuitive understanding of the effect of the size and shape of the resonator on its performances [22]. This model does not address the very complex spurious mode problem. It is however very helpful for the electronic designer to understand the link between the size, or in a more meaningful way the impedance, and the coupling and Q factors.

1.3.5 A/p Empirical Model

In a previous section, we introduced the BVD and MBVD equivalent circuits. These circuits correspond to ideal resonators in the sense that no electrical parasitics are taken into account. In real devices, many electrical parasitics exist, such as capacitance to the substrate, capacitance between pads, capacitance to the ground ring, resistance in the substrate, inductance associated with the connection lines, and so on. Fitting the impedance curve of a real resonator with a simple MBVD equivalent circuit is usually possible. However, the physical meaning of each element is then not straightforward since their value refers to both resonator's intrinsic phenomena and electrical parasitics.

In what follows, the method that was applied is the following. Impedance curves of real resonators were fitted with complex equivalent circuits, based on the MBVD circuit, but with many additional components describing the electrical parasitics. The value of each element corresponding to a parasitic element was calculated automatically from the geometrical and physical data of the device. The validity of the automatic calculation of the parasitic elements was checked with EM simulations. In other words, the only free fitting parameters are the values of the six elements of the MBVD circuit, C_0, R_0, C_m, L_m, R_m, and R_s. As previously explained, the value of R_s is set by EM simulation so that only five free fitting parameters are left. Excellent fits are obtained in most cases, though fitting might be difficult in certain cases, when a myriad of spurious modes is present, for example.

With this method, it is possible to separate contributions from the ideal resonator (MBVD circuit) and from parasitic elements. The model presented hereafter deals with the properties of ideal resonators. The aim of the model is to describe the variation of the properties of the ideal resonator with the size and shape of the device. Figure 1.10 shows typical variations of $k_{eff}^{2\ ideal}$, Q_s^{ideal}, and Q_p^{ideal} extracted

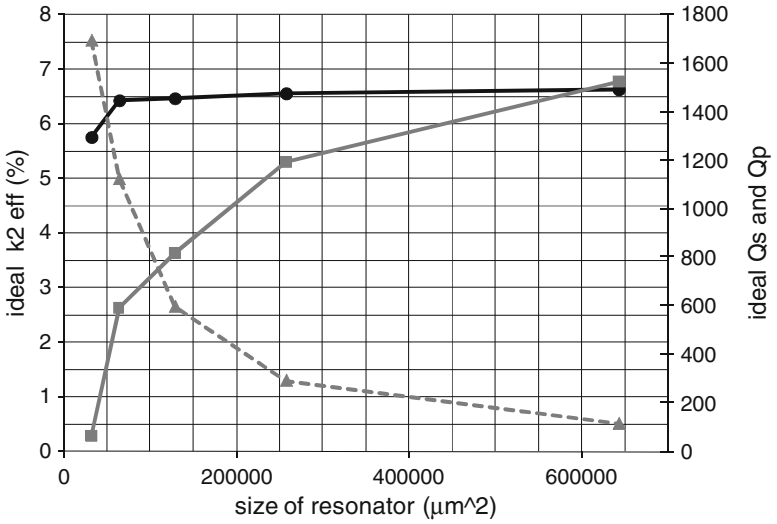

Fig. 1.10 Size dependence of $k_{eff}^{2\ ideal}$ (*black line*), Q_s^{ideal} (*dashed gray line*), and Q_p^{ideal} (*gray line*). These data were obtained by fitting the measured impedance curve of real resonators with a complete equivalent circuit, including parasitics, and then extracting the properties of the ideal resonator

from the fit of measured real resonators. Both $k_{eff}^{2\ ideal}$ and Q_p^{ideal} increase with the size of the resonator while Q_s^{ideal} decreases (by a factor 17 between the smallest and the largest resonators!).

Consider first the coupling coefficient $k_{eff}^{2\ ideal}$. In the model, the resonator is divided in two regions, center and edge. Consider a small volume in the central region at a frequency close to the resonance or antiresonance frequency. Due to the piezoelectric effect, this small volume expands and contracts at the same frequency as the driving electrical field. Since the surrounding material does the same, this small volume vibrates more or less freely. The situation at the edge of the resonator is not the same. The material outside the resonator does not move and prevents the material at the edge of the resonator to move freely. So there is a transition region from the central region of the resonator where a free vibration takes place, to the outside region of the resonator where the material is at rest. In a real resonator, we can assume that this transition occurs smoothly over a certain distance and that part of this transition lays in the resonator itself and part in the outside region. The damping in the transition region of the resonator leads to a lowering of the total coupling coefficient of the resonator. Let's now consider a model where the smooth transition is replaced by an abrupt transition between a region where material moves freely and a region where the material has zero displacement. In other words, there is a region with a coupling coefficient $k_{eff}^{2\ free}$ and a region with 0 coupling. The situation is sketched in Fig. 1.11.

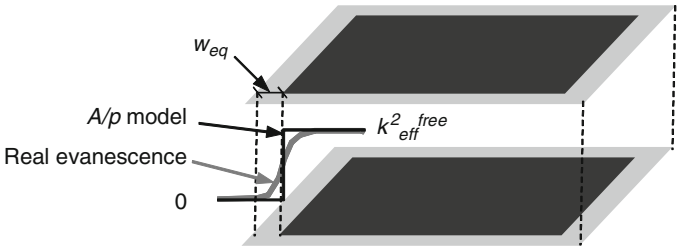

Fig. 1.11 Schematic view of a resonator with a central zone (*dark gray*) and an edge zone (*light gray*). In the central zone, the piezoelectric layer moves freely, and the coupling coefficient is $k_{eff}^{2\ free}$. In the edge zone, the piezoelectric layer is totally clamped, and the coupling coefficient is zero

The width of the region at the edge of the resonator with 0 coupling is chosen so as to produce the same decreasing effect on the average coupling coefficient value $k_{eff}^{2\ ideal}$ of the whole resonator, as the smooth evanescence zone does. It is called equivalent width w_{eq}. It follows that the coupling coefficient of an ideal resonator, $k_{eff}^{2\ ideal}$, is obtained from the volume ratio of the part with $k_{eff}^2 = k_{eff}^{2\ free}$ and the part with $k_{eff}^2 = 0$. The volume with $k_{eff}^2 = 0$ is proportional to the surface given by w_{eq} times the perimeter p of the resonator. It follows

$$k_{eff}^{2\ ideal} = k_{eff}^{2\ free} \frac{A - w_{eq}\, p}{A},$$

or

$$k_{eff}^{2\ ideal} = k_{eff}^{2\ free} \left(1 - \frac{w_{eq}}{A/p}\right), \tag{1.20}$$

where A is the total area of the resonator.

Figure 1.12 shows an example of this model applied in a real case. The resonators under study are circular resonators at 2.4 GHz. The 1-µm AlN layer is sandwiched between a Pt bottom electrode and an Al/W top electrode. The resonators stand over a W-SiO_2 Bragg mirror. Both $k_{eff}^{2\ ideal}$ given by the model and $k_{eff}^{2\ ideal}$ extracted from the measurement fit very well. It is noteworthy that, in accordance with (1.20), $k_{eff}^{2\ ideal}$ depends on A/p, the ratio between surface and perimeter of the resonator, rather than on A solely. The ratio A/p as a function of A depends on the shape of the resonator. For a given A, A/p is larger for a square than for a triangle and is the largest for a circle. Independently of the shape, A/p always increases with increasing A. In other words, small resonators have small A/p. In Fig. 1.12, small resonators show a reduced $k_{eff}^{2\ ideal}$ relative to large ones. A resonator with $A/p = w_{eq}$ is completely blocked and has 0 coupling. In the case of Fig. 1.12, $w_{eq} = 5\ \mu$m. A circular resonator with $A/p = 5\ \mu$m has a diameter of 20 µm. Any circular resonator with a diameter equal or smaller to 20 µm and made with the same

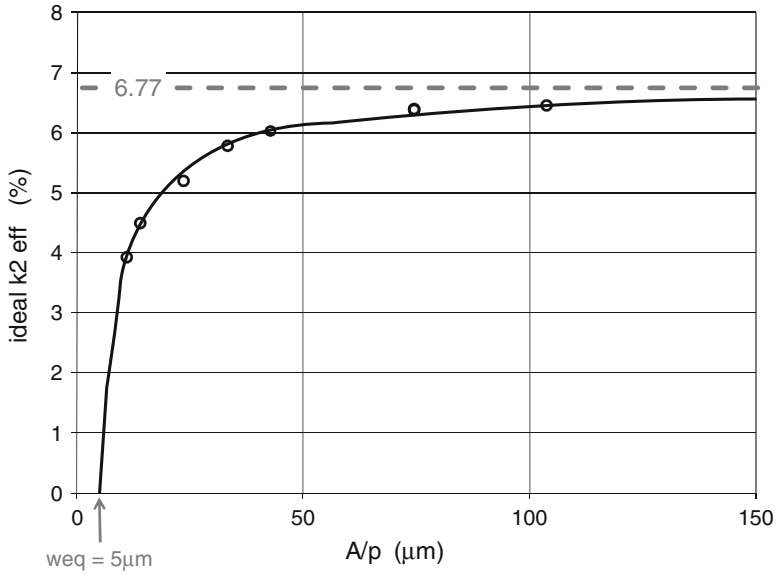

Fig. 1.12 Comparison of $k_{eff}^{2\ ideal}$ extracted from the measurement of seven resonators of different sizes (*circular dots*) with the theoretical curve given by (1.20) (*black line*). The parameters for the curve are $k_{eff}^{2\ free} = 6.77\%$ and $w_{eq} = 5\ \mu$m. ©2008 IEEE. Reprinted, with permission, from [22]

technology as the one used in Fig. 1.12 cannot work. On the other side, an infinitely large resonator would have a coupling corresponding to $k_{eff}^{2\ free}$ (6.77% in the case of Fig. 1.12). The seven resonators shown in Fig. 1.12 have impedances of 500, 300, 100, 50, 30, 10, and 5 Ω at 2.4 GHz, the largest one (5 Ω) being on the right of the curve.

Let's now consider the case of the parallel Q factor, Q_p^{ideal}. Assuming that at the antiresonance the losses are mainly acoustic [20, 23], Q_p^{ideal} can be modeled as follows. On one hand, part of the losses is proportional to the volume of the resonator, which is proportional to the surface A of the resonator. For example, energy leaks through the mirror or is internally dissipated (viscous damping, scattering at inhomogeneities...). On the other hand, part of the energy leaks laterally at the edge of the resonator (lateral waves, friction...). This energy leakage is proportional to the lateral surface delimiting the resonator, which is proportional to the perimeter p. Finally, the total energy injected in the resonator is proportional to the area A. The Q factor being defined as the ratio between the total energy to the dissipated energy, Q_p^{ideal} can be written as

$$Q_p^{ideal} = \frac{A}{aA + bp},$$

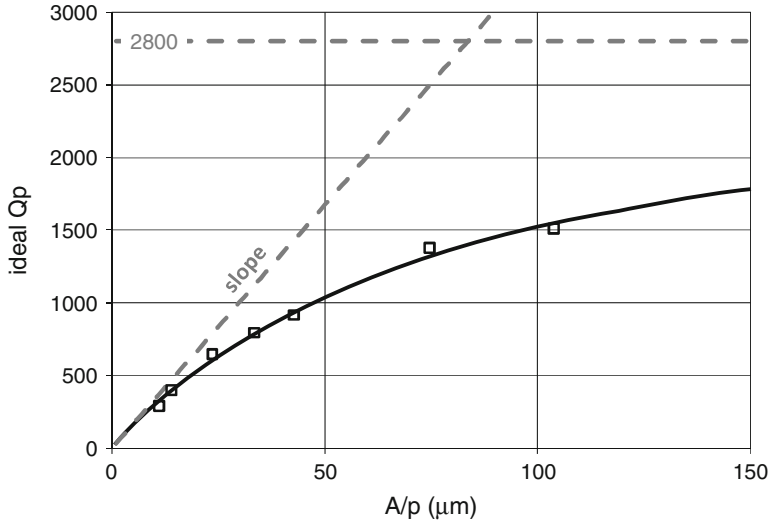

Fig. 1.13 Comparison of Q_p^{ideal} extracted from the measurement of seven resonators of different sizes (*square dots*) with the theoretical curve given by (1.22) (*black line*). The parameters for the curve are $Q_p^{free} = 2800$ and $slope = 33.5\ \mu\text{m}^{-1}$. ©2008 IEEE. Reprinted, with permission, from [22]

or

$$Q_p^{ideal} = \frac{1}{a + \frac{b}{A/p}}, \tag{1.21}$$

where a and b are proportionality constants. Identifying $a = 1/Q_p^{free}$ and renaming $b = 1/slope$, it follows

$$Q_p^{ideal} = \frac{1}{\frac{1}{Q_p^{free}} + \frac{1}{slope\ (A/p)}}. \tag{1.22}$$

The unit of $slope$ is μm^{-1}. As in the case of the coupling, there is an A/p dependence. In Fig. 1.13, the model for Q_p^{ideal} is applied to the same seven resonators that were used with $k_{eff}^{2\ ideal}$ in Fig. 1.12. Again, Q_p^{ideal} given by the model and Q_p^{ideal} values extracted from the measurement fit very well. Small resonators have a Q_p that tends to zero while $Q_p^{free} = 2{,}800$ gives an upper limit to the parallel Q factor that can be achieved with this technology. Whether Q_p^{free} is limited by the mirror, internal losses, or other mechanisms is out of the scope of this chapter.

The case of the series Q factor is more complicated to handle. Let's first recall that the model presented up to now deals with ideal resonators, free of any parasitics. In that sense, the series resistance R_s of the interconnections and electrodes is also a parasitic element. Hence, this model will deal with Q_s^{intr}, the intrinsic Q_s factor

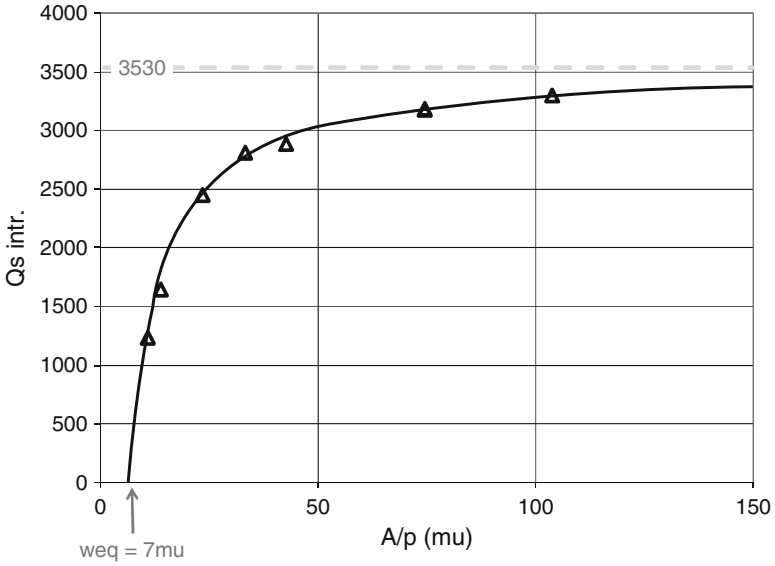

Fig. 1.14 Comparison of Q_s^{intr} extracted from the measurement of seven resonators of different sizes (*triangular dots*) with the theoretical curve given by (1.23) (*black line*). The parameters for the curve are $Q_s^{intr\,free} = 3{,}530$ and $w_{eq} = 7\ \mu$m. ©2008 IEEE. Reprinted, with permission, from [22]

described previously. Knowing the intrinsic Q_s factor presents some interest in itself. However, as we will see, the model also allows understanding Q_s of real resonators, which is of much greater interest.

One way to reason would be to treat the series resonance similarly to the parallel resonance. However, the behaviors in the two cases are quite different. While acoustic losses dominate at the antiresonance, ohmic losses dominate at the resonance [20, 23]. Hence, a dependence of Q_s on A/p similar to that of Q_p is questionable. Furthermore, the extraction of the intrinsic Q_s from measured data rather suggests a law similar to that of k_{eff}^2. By analogy with (1.20), we get

$$Q_s^{intr} = Q_s^{intr\,free}\left(1 - \frac{w_{eq}}{A/p}\right). \tag{1.23}$$

With this model, the resonator is again divided into two regions: a central region where it is free to move and for which $Q_s^{intr} = Q_s^{intr\,free}$ and a blocked edge region where all the energy is dissipated and hence leading to $Q_s^{intr} = 0$. Figure 1.14 shows the results obtained for Q_s^{intr} of the seven resonators presented previously. Again, the model and the values for Q_s^{intr} of the resonators extracted from the measurement fit very well. Small resonators with $A/p < w_{eq}$ have a Q_s^{intr} of 0. It increases then rapidly with the size of the resonator. Large resonators present a Q_s^{intr} close to the upper limit given by $Q_s^{intr\,free} = 3{,}530$.

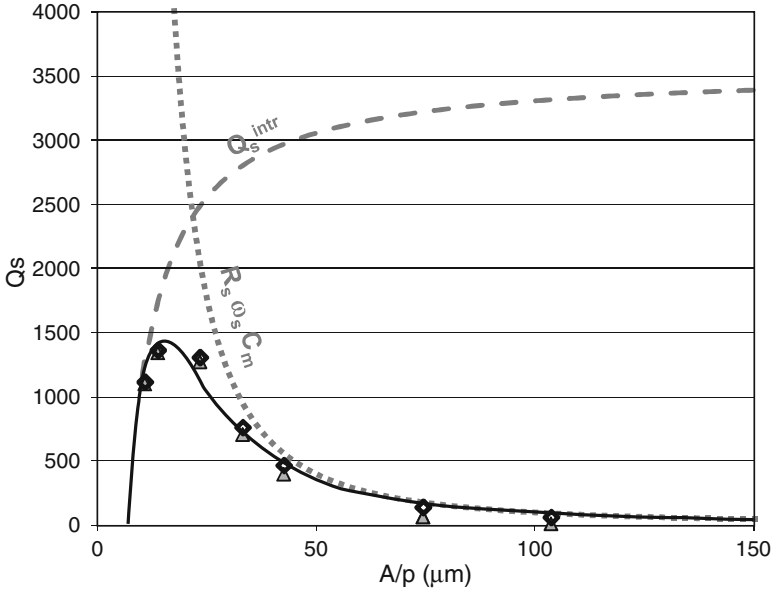

Fig. 1.15 Comparison of Q_s^{ideal} extracted from the measurement of seven resonators of different sizes (*black diamond dots*) with the theoretical curve given by (1.24) (*black line*). The curve is asymptotically limited by two curves. On the small resonators side, it is limited by Q_s^{intr} (*dashed gray line*) (see Fig. 1.14). This is the first term in (1.24) denominator. On the large resonators side, it is limited by the $R_s \cdot \omega_s \cdot C_m$ term in (1.24) (*dotted gray line*). Real Q_s are also shown (*gray triangle dots*). ©2008 IEEE. Reprinted, with permission, from [22]

Knowing the intrinsic Q_s factor Q_s^{intr} and the ideal coupling coefficient $k_{eff}^{2\ ideal}$, we are now in a position to calculate Q_s^{ideal}, which includes the ohmic losses. Using the following definitions,

$$Q_s^{intr} = \frac{L_m\,\omega_s}{R_m} = \frac{1}{C_m\,\omega_s\,R_m} \quad \text{and} \quad Q_s^{ideal} = \frac{1}{C_m\,\omega_s\,(R_m + R_S)},$$

it follows

$$Q_s^{ideal} = \frac{1}{\frac{1}{Q_s^{intr}} + C_m\,\omega_s\,R_S}, \tag{1.24}$$

where Q_s^{intr} is a function of A/p according to (1.23) and C_m is also a function of A/p since C_m depends on k_{eff}^2 according to $C_m = (C_0 \cdot k_{eff}^2 \cdot 8/\pi^2)/(1 - C_0 \cdot k_{eff}^2 \cdot 8/\pi^2)$, which is a function of A/p according to (1.20).

Figure 1.15 shows the results obtained for Q_s^{ideal} of the seven resonators presented previously. Again, the model and the values for Q_s^{ideal} factor of the resonator, extracted from the measurement, fit very well. It can be seen that Q_s^{ideal} of small resonators is limited by Q_s^{intr}. Q_s^{ideal} reaches then a maximum and then begins to decreases severely for larger resonators. This is due to the $R_s \cdot \omega_s \cdot C_m$ term in (1.24).

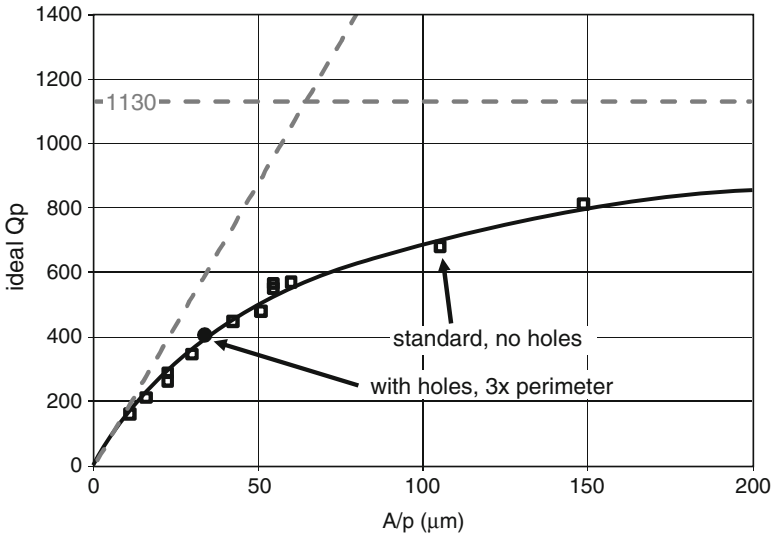

Fig. 1.16 Q_p^{ideal} of standard resonators are shown with *square dots*. The *full circular dot* at $A/p = 34$ is a resonator with the same surface A as the standard resonator with $A/p = 105$, but holes in the top electrode increase the perimeter p by a factor three. ©2008 IEEE. Reprinted, with permission, from [22]

Large resonators have small impedances. Since R_s does not vary much with the size of the resonators, it is logical that it has more impact on low-impedance resonators. Figure 1.15 also shows the real Q_s, extracted directly from the measurements. The difference between the real Q_s and Q_s^{ideal} is due to the parasitics. Again, it is logical that they impact more low-impedance resonators. However, the trend for Q_s of real resonators is similar to that of ideal resonators.

The survey of this model will end up with three examples. The first one is illustrated by Fig. 1.16. All resonators but one were standard rectangular resonators of different sizes. The extraction of their ideal coupling and Q factors from the measured impedance was performed. The two asymptotes of the Q_p^{ideal} curve are defined by $Q_p^{free} = 1{,}130$ and a slope of $17.6\,\mu\text{m}^{-1}$. These resonators were fabricated with an older technology, based on an AlN-SiO_2 Bragg mirror and Al top electrode, and without any frame on the edge of the resonators. The special resonator is a resonator with holes in the top electrode. The surface covered with top metal is equal to that of the standard resonator with $A/p = 105\ \mu$m. Due to the holes, the length of edges on this resonator was multiplied by three relative to the perimeter of the standard resonator, resulting in $A/p = 34$. As shown in Fig. 1.16, Q_p^{ideal} drops from 680 without holes to 400 with holes, as predicted by the model.

The second example also shows the impact of the perimeter, but this time by varying the shape. Figure 1.17 shows Q_p^{ideal} for resonators with the same surface but different shapes. The triangular resonator has the longest perimeter and consequently the smallest A/p ratio. The circular resonator has the largest A/p ratio. As shown on Fig. 1.17, Q_p^{ideal} increases with decreasing perimeter.

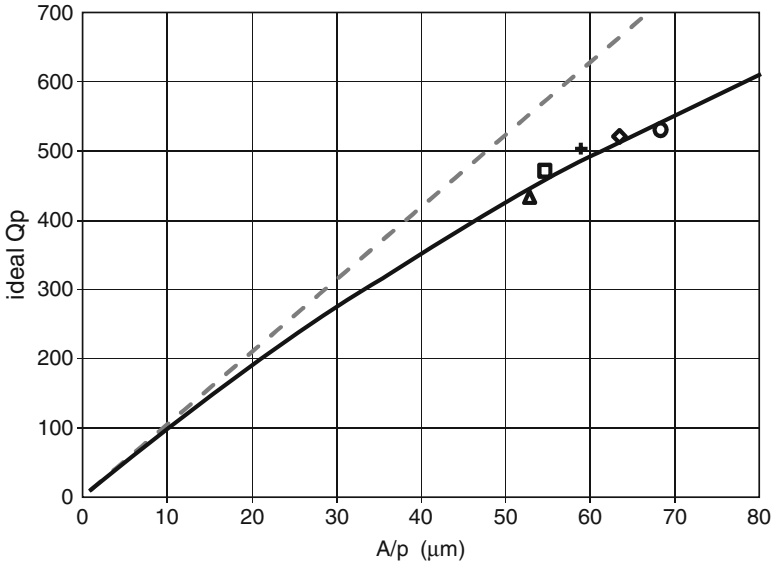

Fig. 1.17 Q_p^{ideal} vs A/p for resonators with the same surface A, but with different shapes. The shapes with increasing A/p are *triangle, rectangle, ellipse, pentagon, and circle*. ©2008 IEEE. Reprinted, with permission, from [22]

The last example shows how the model can be used successfully in the design of filters. The equivalent circuit of the ladder filter is built in the same way as for resonators. The resonators are represented with an MBVD circuit with values corresponding to ideal resonators. The necessary elements are then added to model the parasitics. Their value is calculated automatically from geometrical and material data. Figure 1.18 shows the very good agreement between simulation and measured data. The insertion loss (IL), the notches, the rejection, and VSWR fit very well. This 50-MHz bandwidth ladder filter was designed for the 2.4–2.48-GHz ISM band. It comes from the same wafer as the seven resonators shown in Figs. 1.12–1.15. Its minimum IL is very good at 0.98 dB. The maximum insertion loss in the 50-MHz band is 1.6 dB. The minimum rejection is 20 dB in the 1.3–1.8-GHz range. There is a nice rejection notch between 6 and 7 GHz. No external components were added to this filter to enhance the performances.

1.4 Conclusion

After two decades of technology development, the first thin-film BAW resonators have hit the market as passive elements in front-end filters and duplexers. Beside this mainstream application, these components can be advantageously used in many

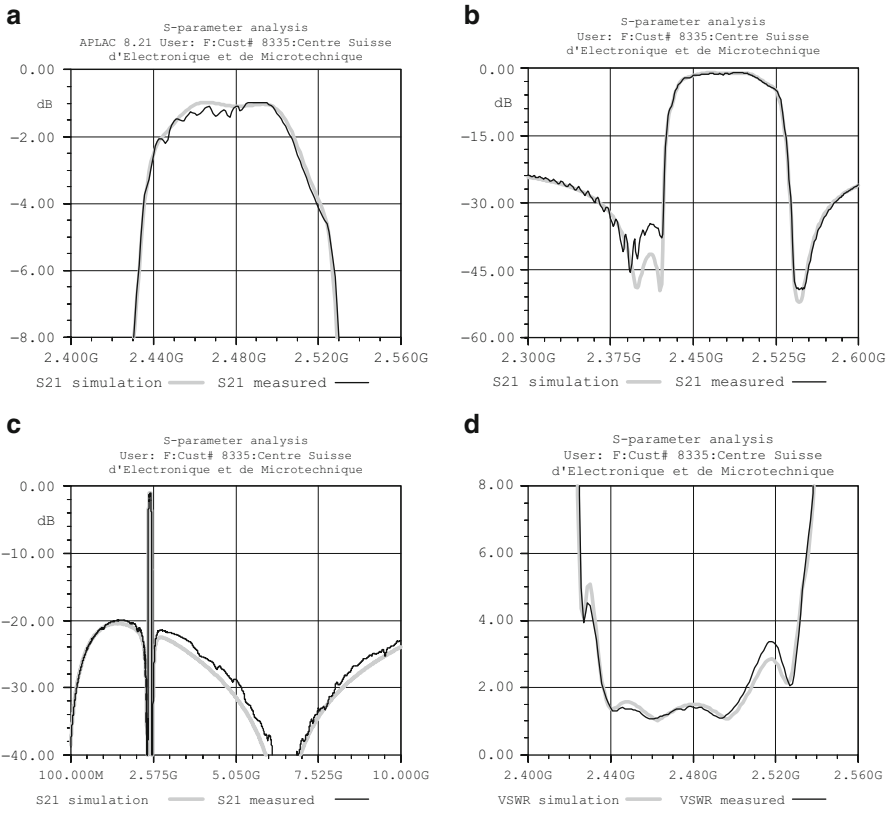

Fig. 1.18 Simulation (*gray*) vs measurement (*black*) for a ladder filter: (**a**) Insertion loss, (**b**) notches, (**c**) rejection, and (**d**) VSWR. ©2008 IEEE. Reprinted, with permission, from [22]

other communication systems, owing to their very high performances. For example, very low power transceivers could benefit from the high quality factors of these resonators. In any case, whatever the application, the designer needs tools for describing the behavior of these piezoelectric devices. Models are necessary for the design of complex filters made from the resonators but also for the simulation of electronic circuits that include BAW devices. The empirical model proposed in this chapter enables the accurate description of the resonators, by computing the value of the parasitic elements that must be added to the simple electrical model used usually for piezoelectric resonators such as quartz crystals. Indeed, in the case of thin-film resonators, these elements have a major influence on the performance of the devices in terms of Q factor and coupling coefficient and must hence be accounted for. The fact that the values of the parasitic elements depend on the size of the devices is implemented in the model by evaluating the properties of the resonators as a function of their perimeter and area. The applicability of this model has been

shown by the excellent match between calculated and measured parameters, both for resonators and more complex filters. Now that the fabrication technology and the modeling tools have come to a quite mature stage, and even though additional improvements could still be beneficial, the challenge can be moved towards the design of new communication systems. Indeed, the integration of BAW resonators with tailored properties in advanced circuit architectures will certainly open new opportunities.

References

1. T.R. Sliker, D.A. Roberts, J. Appl. Phys. **38**, 2350 (1967)
2. T.W. Grudkowski, J.F. Black, T.M. Reeder, et al., Appl. Phys. Lett. **37**, 993 (1980)
3. K.M. Lakin, J.S. Wang, Appl. Phys. Lett. **38**, 125 (1981)
4. K. Nakamura, H. Sasaki, H. Shimizu, Electron. Lett. **17**, 507 (1981)
5. K.M. Lakin, J.S. Wang, G.R. Kline, et al., in *IEEE 1982 Ultrasonics Symposium*, 1982, pp. 466–475
6. H. Satoh, Y. Ebata, H. Suzuki, C. Narahara, in *IEEE 39th Frequency Control Symposium*, 1985, pp. 361–366
7. K.M. Lakin, K.T. McCarron, R.E. Rose, in *IEEE 1995 Ultrasonics Symposium*, 1995, pp. 905–908
8. R.C. Ruby, P. Bradley, Y. Oshmyansky, A. Chien, in *IEEE 2001 Ultrasonics Symposium*, Münich, 2001, pp. 813–821
9. R. Aigner, J. Kaitila, J. Ellä, et al., in *IEEE 2003 MTT Symposium*, 2003, pp. 2001–2004
10. R. Lanz, C. Lambert, L. Senn, et al., in *IEEE 2006 Ultrasonics Symposium*, Vancouver, Canada, 2006, pp. 1481–1485
11. M.A. Dubois, J.F. Carpentier, P. Vincent, C. Billard, G. Parat, C. Muller, P. Ancey, P. Conti, IEEE J. Solid State Circ. **41**(1), 7 (2006)
12. B.A. Auld, *Acoustic Fields and Waves in Solids* (Robert E. Krieger Publishing Company, Malabar, 1973)
13. W.P. Mason, *Physical Acoustics* (Academic, New York, 1964)
14. D. Royer, E. Dieulesaint, *Ondes élastiques Dans Les Solides* (Masson, Paris, 1996)
15. J. Kaitila, M. Ylilammi, J. Ellä, R. Aigner, in *IEEE 2003 Ultrasonics Symposium*, Hawaii, 2003, pp. 84–87
16. G.G. Fattinger, S. Marksteiner, J. Kaitila, R. Aigner, in *IEEE 2005 Ultrasonics Symposium*, Rotterdam, 2005, pp. 1175–1178
17. J.D. Larson, P.D. Bradley, S. Wartenberg, R.C. Ruby, in *IEEE 2000 Ultrasonics Symposium*, Puerto Rico, 2000, pp. 863–868
18. T. Makkonen, A. Holappa, J. Ellä, M.M. Saloma, in *IEEE 2001 Ultrasonics Symposium*, Münich, 2001, pp. 1241–1258
19. K.M. Lakin, K.G. Lakin, in *IEEE 2003 Ultrasonics Symposium*, Hawaii, 2003, pp. 74–79
20. R.F. Milsom, H.P. Löbl, C. Metzmacher, in *Proceedings 2nd International Symposium on Acoustic Wave Devices for Future Mobile Communication Systems*, Chiba, 2004, pp. 143–154
21. R.C. Ruby, J.D. Larson, R.S. Fazzio, C. Feng, in *IEEE 2005 Ultrasonics Symposium*, Rotterdam, 2005, pp. 1832–1835
22. C. Muller, M.A. Dubois, in *IEEE 2008 Ultrasonics Symposium*, Beijing, 2008, pp. 1552–1556
23. R. Thalhammer, R. Aigner, in *IEEE 2005 Ultrasonics Symposium*, Rotterdam, 2005, pp. 225–228

Chapter 2
Contour-Mode Aluminum Nitride Piezoelectric MEMS Resonators and Filters

Gianluca Piazza

Abstract This chapter describes the aluminum nitride MEMS contour-mode resonator (CMR) technology and its application to RF filtering and frequency synthesis. The CMR technology is a new class of piezoelectric resonant devices that has the ability to span a broad range of frequencies from few MHz up to GHz on the same silicon chip and attain motional resistances in the range of 25–250 Ω and quality factors above 1,000 in air over the entire frequency spectrum. These laterally vibrating AlN microstructures not only provide the advantages of compact size, low power consumption, and compatibility with high yield mass producible components but will also enable paradigm-shifting solutions for reconfigurable RF front ends and simpler frequency synthesizers. In this chapter, basic analytical design procedures are presented to explain the principle of operation of one and two port AlN CMRs and their use in oscillator circuits. Fundamentals of microfabrication techniques employed for making AlN CMR are briefly introduced. Different methods for arraying these devices and form either electrically or mechanically coupled filters are described. Key device parameters that affect filter insertion loss, bandwidth, and rejection are highlighted for the different kind of configurations. Finally, potential applications of these devices in next-generation cognitive radios and future research directions are presented.

2.1 Aluminum Nitride MEMS Contour-Mode Resonator Technology

Recent advances in surface micromachining techniques have enabled the realization of miniaturized and high-quality factor acoustic resonators that can be integrated

G. Piazza (✉)
Department of Electrical and Computer Engineering, Carnegie Mellon University, Roberts Engineering Hall, Room 333, USA
e-mail: piazza@ece.cmu.edu

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_2,

with state-of-the-art CMOS electronics [1–7]. Among these MEMS devices, a new class of resonators, dubbed contour-mode (contour-mode resonators, CMR) because of their in-plane mode of vibration, has received large attention from the research community. In fact, these resonant devices offer the unique capability of providing multiple frequencies of operation on the same silicon substrate. Electrostatic and piezoelectric transduction mechanisms have emerged as the preeminent techniques for driving and sensing resonant vibrations in micromechanical structures. Electrostatically transduced resonators have demonstrated high-quality factors and high frequencies (up to GHz by employing overtones) but suffer from large motional impedances that make their interface with 50 Ω RF systems troublesome. Instead, piezoelectric actuation of aluminum nitride CMRs addresses this fundamental challenge and offers low motional resistance (25–250 Ω) while maintaining high-quality factors (1,000–4,000) and capacitance values (0.2–1 pF) that ease their interface with state-of-the-art circuitry.

Contour modes of vibration can be excited in c-axis-oriented AlN films via the equivalent d_{31} piezoelectric coefficient. By applying an electric field across the film sandwiched between a top and bottom electrode, the MEMS structure expands laterally and can be excited in resonant vibrations whose frequency is set by one of the in-plane dimensions of the device.

The most promising structures demonstrated to obtain high Q (1,000–4,000) and high frequency of operations (10 MHz–1.6 GHz) are rings and rectangular plates as shown in Fig. 2.1. The frequency of vibration, f_o, is generally set by the width of the structure, W (Fig. 2.1), whereas the second dimension can be employed to control the equivalent motional resistance, R_M, and static capacitance, C_O, of the device. Either fundamental or higher order mode can be selectively excited in rectangular or annular plates (Fig. 2.1) by properly patterning the top electrode and therefore defining the effective width of the resonant element. The use of higher order modes of vibration is deemed necessary in order to reduce the sensitivity to lithographic errors and misalignments for devices operating at frequencies above 400 MHz, for which the definition of small features would otherwise significantly complicate the ability to accurately set the frequency. The multielectrode configuration (used to selectively excite higher order modes) reduces the sensitivity to lithographic tolerances by relying on the definition of a precise pitch rather than an absolute dimension, as in the case of the fundamental mode device. In addition, the frequency defining element becomes the electrode (rather than the AlN films), which is thin (50–200 nm) and easier to pattern into very small features. A fundamental advantage of this resonator technology over film bulk acoustic resonators (FBARs) or surface acoustic wave (SAW) resonators is that the device center frequency and electromechanical coupling can be set exclusively at the CAD layout level, therefore enabling multiple frequencies of vibration on the same silicon chip and greatly reducing manufacturing tolerances to film thicknesses.

In general, the equivalent circuit representation for a piezoelectric transducer (such as a contour-mode device) is a simplified version of the Mason's model as shown in Fig. 2.2.

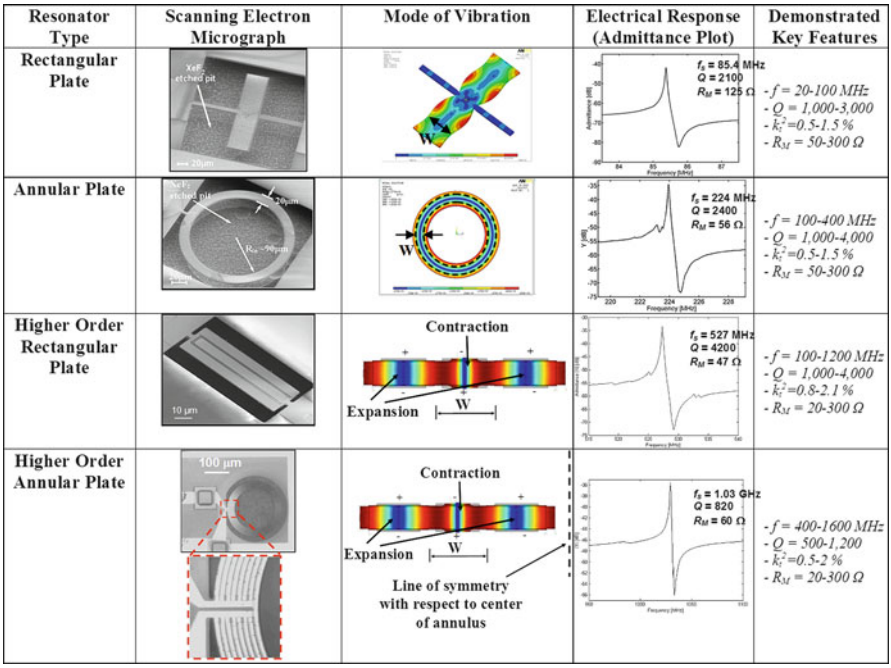

Fig. 2.1 Summary of most significant AlN contour-mode resonator topologies demonstrated to date. Reprinted with permission from [8]. ©2009, American Vacuum Society

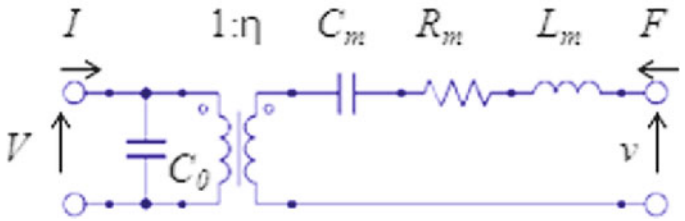

Fig. 2.2 Conventional Mason lumped circuit model for a piezoelectric transducer. Reprinted from [9], ©2007 with permission from Elsevier

As shown in Fig. 2.2, the transducer can be modeled by an intrinsic capacitance, C_O, representing the physical capacitance of the electroded part of the piezoelectric device; a transformer, whose turn ratio, η, represents the conversion between electrical and mechanical variables at a specific location of the device (generally the point of maximum displacement); and motional capacitance, C_m, resistance R_m, and inductance L_m, representing the mechanical variables of the MEMS resonator, being associated with its compliance $1/k_{eq}$, damping c_{eq}, and mass m_{eq}, respectively (see Table 2.2):

$$C_0 = \varepsilon_{33} \frac{\text{Electroded area}}{\text{Thickness of electroded area}}, \tag{2.1}$$

Table 2.1 Correspondence between mechanical and electrical variable used to describe mechanical resonators in the two domains

Mechanical variable	Electrical analogue
Force (F)	Voltage (V)
Velocity (v)	Current (I)
Compliance ($1/k_{eq}$)	Capacitance (C_m)
Damping (c_{eq})	Resistance (R_m)
Mass (m_{eq})	Inductance (L_m)

$$\eta = \frac{F}{V} = \frac{I}{v}, \qquad C_m = \frac{1}{k_{eq}}, \qquad R_m = c_{eq}, \qquad L_m = m_{eq},$$

where ε_{33}, is the permittivity of the piezoelectric material in the thickness direction and F, V, I, and v are defined in Table 2.1. The equivalent lumped mechanical variables can be calculated according to the techniques that are explained in detail, for example, in Johnson [10] or in Tilmans [11, 12].

The AlN plate can be excited into vibrations by two means of applying an electric field across the film thickness. The two most successful implementations to date are based on thickness field excitation (TFE) and lateral field excitation (LFE) (Fig. 2.3). LFE results in a more robust implementation since it is less sensitive to electrode alignment but suffers from reduced electromechanical coupling and effective device reactance, which simultaneously make the resonator interface with external circuitry more complicated. For these reasons, TFE is the preferred embodiment, and the one that will be discussed in further details in the following sections for both one-port and two-port implementations.

2.1.1 One-Port AlN Contour-Mode Resonators

A one-port resonator is a device for which the simplified model of Fig. 2.2 is terminated in a short, representing, in the acoustic domain, a perfectly reflective boundary such as air. In one-port configurations, the static capacitance of the device, C_O, appears in parallel with the motional components of the resonator and its value depends, if fringing fields and piezoelectric stiffening effects are neglected, only on the electroded area. The equivalent electrical model can be reduced to the conventional Butterworth Van Dyke (BVD), generally modified to include the effect of parasitic components (R_S and R_O), which play a significant role at the MEMS scale (Fig. 2.4). The parasitic element, R_S, represents the equivalent series resistance due to the resonator electrodes and via (dominant term) and AlN films. R_S and R_O are generally extracted from experimental measurements. The other parameters, R_M, L_M, and C_M, represent the equivalent motional parameters of the resonator in the electrical domain and are given by (for a rectangular geometry)

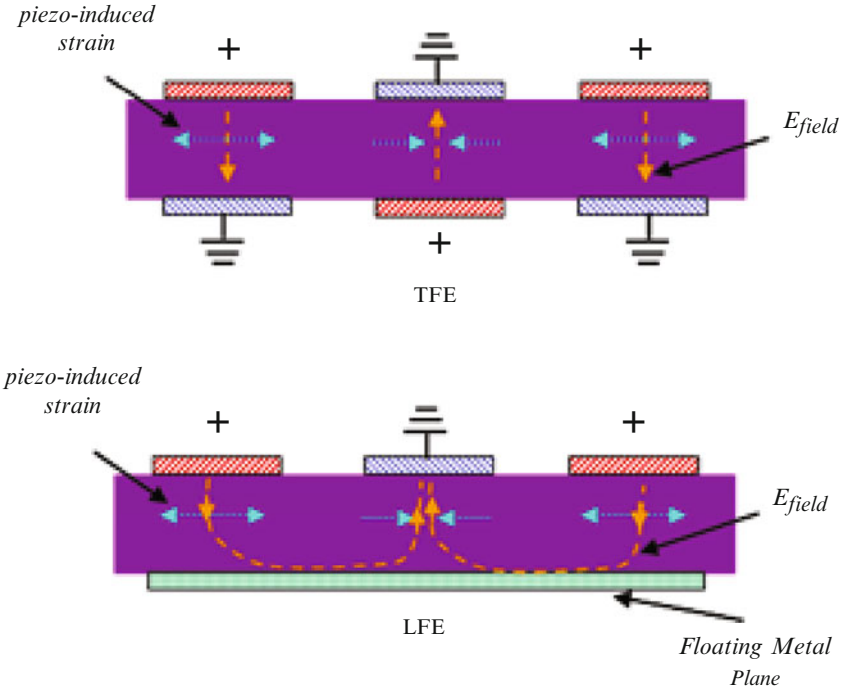

Fig. 2.3 Two possible schemes of excitation of lateral modes of vibrations in AlN piezoelectric films. Electric field across the film thickness (which induces lateral strain in the piezoelectric film via the d_{31} coefficient) can be applied either via a thickness field excitation (TFE, top of figure) or a lateral field excitation (LFE, bottom of figure). The bottom floating electrode shown in the LFE embodiment is not necessary but helps in attaining a better confinement of the electric field and a correspondingly higher electromechanical coupling. Reprinted from [9], ©2007 with permission from Elsevier

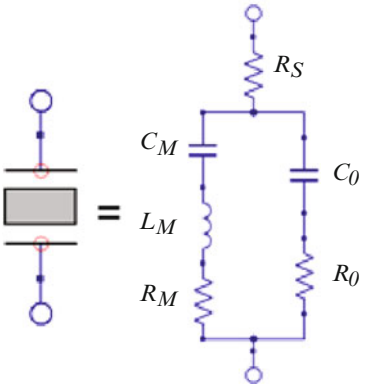

Fig. 2.4 Circuital symbol for an electromechanical resonator and equivalent MBVD model used to represent its behavior in the electrical domain. ©2008 IEEE. Reprinted, with permission, from [13]

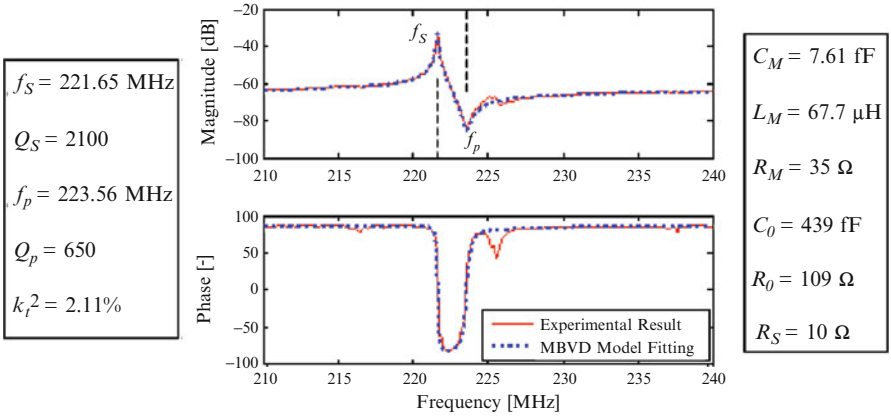

Fig. 2.5 Measured admittance plot and MBVD model fitting for a piezoelectric AlN contour-mode resonator at 221.6 MHz

$$R_M = \frac{1}{n}\frac{\pi}{8}\frac{T}{L}\frac{\rho_{eq}^{1/2}}{E_{eq}^{3/2}d_{31}^2 Q} = \frac{\pi^2}{8}\frac{1}{\omega_s C_O}\frac{1}{k_t^2 Q}, \tag{2.2}$$

$$L_M = \frac{1}{n}\frac{WT}{8L}\frac{\rho_{eq}}{E_{eq}^2 d_{31}^2} = \frac{\pi^2}{8}\frac{1}{\omega_s^2 C_O}\frac{1}{k_t^2}, \tag{2.3}$$

$$C_M = n\frac{8}{\pi^2}\frac{WL}{T}E_{eq}d_{31}^2 = \frac{8}{\pi^2}C_O k_t^2, \tag{2.4}$$

where ω_s is the device center frequency:

$$\omega_s = \frac{\pi}{W}\sqrt{\frac{E_{eq}}{\rho_{eq}}}. \tag{2.5}$$

L, W, and T refer to the length, width, and thickness of each resonator or individual finger forming the AlN plate (as in Fig. 2.1); n is the number of sub-resonators (fingers); E_{eq} and ρ_{eq} are the equivalent in-plane modulus of elasticity and mass density of AlN and the stacked electrodes; d_{31} is the equivalent (3,1) piezoelectric coefficient of AlN; and k_t^2 is the equivalent (3,1) electromechanical coupling coefficient for AlN. In the case of annular geometry, (2.2)–(2.5) still hold, but the resonator length, L, needs to be substituted by $2\pi R_{AVE}$, where R_{AVE} constitutes the average radius of the annulus.

For example, Fig. 2.5 shows how accurately the modified BVD (MBVD) model describes the measured admittance response of a 221.6 MHz AlN CMR having $n = 3$, $L = 200\,\mu$m, $T = 2\,\mu$m, and $W = 20\,\mu$m (such as the one shown in Fig. 2.1). In this specific case, for which 200-nm-thick Pt electrodes were used, $E_{eq} = 399$ GPa and $\rho_{eq} = 5,076\,\text{kg/m}^3$, which closely correspond to the bulk values for the AlN/Pt stack.

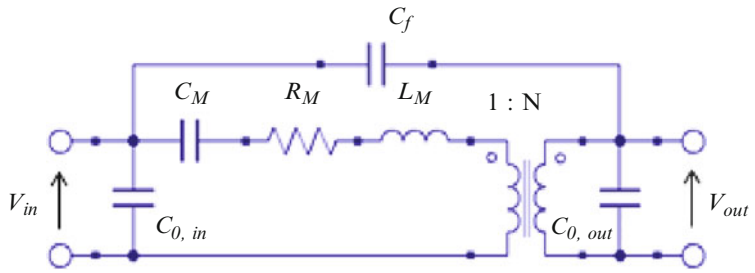

Fig. 2.6 Circuit model for a two-port resonator. ©2007 IEEE. Reprinted, with permission, from [14]

2.1.2 Two-Port Contour-Mode Resonators

A two-port resonator can be represented by two cascaded networks of the type depicted in Fig. 2.2. Ideally, it can be modeled as two resonators coupled at an infinitely stiff location. This implementation is generally encountered in macroscale resonant piezoelectric transformers. The reduced equivalent electrical circuit is shown in Fig. 2.6, in which the parasitic components, R_S and R_O, of Fig. 2.4 have been neglected and a parasitic capacitance, C_f, has been included to account for substrate and device feedthrough between the input and output ports of the resonator.

The main components in the circuit are described by the following equations (for a rectangular geometry and assuming that width, W, length, L, and thickness, T, of each of the elements forming the input and output ports are the same as in Fig. 2.7):

$$C_{O,in} = n_{in}\varepsilon_{33}\varepsilon_O \frac{WL}{T}, \tag{2.6}$$

$$C_{O,out} = n_{out}\varepsilon_{33}\varepsilon_O \frac{WL}{T}, \tag{2.7}$$

$$R_M = \frac{\pi}{8}\frac{T}{L}\frac{\rho_{eq}^{1/2}}{E_{eq}^{3/2}d_{31}^2 Q}\frac{n_{in}+n_{out}}{n_{in}^2} = \frac{\pi^2}{8}\frac{1}{\omega_s k_t^2 Q}\frac{n_{in}+n_{out}}{n_{in}}\frac{1}{C_{O,in}}, \tag{2.8}$$

$$L_M = \frac{WT}{8L}\frac{\rho_{eq}}{E_{eq}^2 d_{31}^2}\frac{n_{in}+n_{out}}{n_{in}^2} = \frac{\pi^2}{8}\frac{1}{\omega_s^2}\frac{1}{k_t^2}\frac{n_{in}+n_{out}}{n_{in}}\frac{1}{C_{O,in}}, \tag{2.9}$$

$$C_M = \frac{8}{\pi^2}\frac{WL}{T}E_{eq}d_{31}^2\frac{n_{in}^2}{n_{in}+n_{out}} = \frac{8}{\pi^2}k_t^2\frac{n_{in}}{n_{in}+n_{out}}C_{O,in}, \tag{2.10}$$

$$N = \frac{\eta_{in}}{\eta_{out}}, \tag{2.11}$$

where the variables are the same as in (2.2)–(2.5) and the subscript *in* and *out* refer to quantities associated either with the input or output ports of the resonator.

A clear advantage of this solution over a one port counterpart is the electrical isolation of the input and output terminals that facilitates the transduction of the

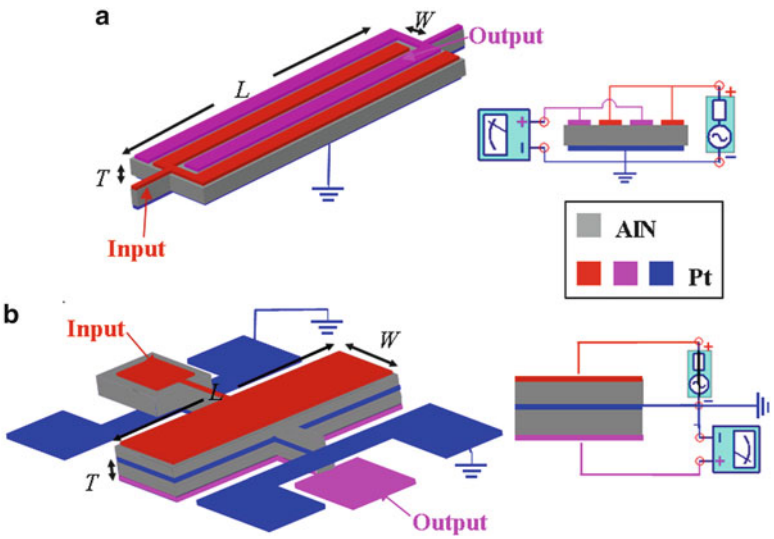

Fig. 2.7 Two of the possible demonstrated embodiment for two-port resonators: (**a**) the two ports are placed parallel to each other in the plane of the resonator and (**b**) the two ports are stacked on top of each other in the vertical direction

resonators at higher frequencies. The two-port design also simplifies the making of electrically coupled filters as shown in Sect. 2.3.1. The disadvantage is that, in a planar solution, the two-port topology increases the value of the motional impedance by 4X for a given transduction area. Two possible embodiments for two-port piezoelectric resonators have already been demonstrated [9, 15] (Fig. 2.7). The two-port configuration can be realized by placing the two transducers adjacent to each others in the lateral direction (Fig. 2.7a) or by stacking the transducers on top of one another (Fig. 2.7b). The second implementation is more effective from the perspective of maximizing the device electromechanical coupling coefficient per unit area. In fact, the device motional resistance matches the one of a one-port device for a given layout area and improves the mechanical coupling into the fundamental mode by effectively driving the device over the entire top surface. Furthermore, the stacked solution limits the input to output capacitive feedthrough by reducing the coupling through the AlN film that is present between the driving and sensing electrodes in the first configuration. The two-port stacked topology was implemented for both rectangular and ring geometries.

2.1.3 Figures of Merit for AlN Contour-Mode Resonators

By examining (2.2)–(2.5) and (2.6)–(2.11), it is evident that all resonator parameters can be defined as a function of the device center frequency, ω_s, static capacitance,

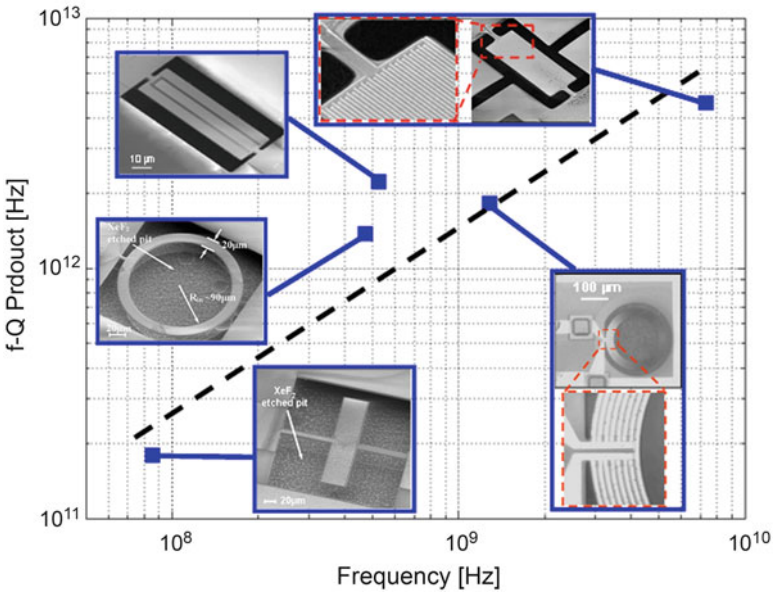

Fig. 2.8 Best fQ product achieved for different CMR devices covering a frequency range from 50 MHz up to 8.5 GHz. Reprinted with permission from [8]. ©2009, American Vacuum Society

C_O, electromechanical coupling, k_t^2, and quality factor, Q. This is the reason why, historically, the two fundamental figures of merit for resonators have always been the $k_t^2 Q$ and fQ products. $k_t^2 Q$ directly affects insertion loss and bandwidth of filters, and power consumption and phase noise in oscillators. fQ is inversely proportional to the noise equivalent mass resolution in a resonator and it is therefore important in defining frequency stability and phase noise in oscillators. In general, the fQ product of most macro- and micromechanical resonators developed to date has been constant. As shown in Fig. 2.8, the CMR technology has instead achieved an $f\,Q$ product that has been increasing with frequency. This trend confirms that it is advantageous to use this technology at very high frequencies but also shows that further improvements in the Q of these devices at lower frequencies should be possible. Q improvement in CMR devices is a subject of current research.

The k_t^2 coefficient describes the conversion of electrical energy into mechanical, and its value depends on the piezoelectric coefficients of AlN, the material stiffness and dielectric constant, the mode of vibration of the resonator, and the corresponding shape of the electrodes used for applying a voltage across the piezoelectric layer. k_t^2 between 0.5% and 2.5% have been demonstrated for the CMR technology, although coefficients as high as 3.5% should be theoretically possible.

Table 2.2 Summary of key properties of AlN films deposited by sputtering techniques. Reprinted from [9], ©2007 with permission from Elsevier

AlN property	Value
d_{31} piezo-coefficient	-2 pC/N
Rocking curve (FWHM)	0.7–2%
Resistivity	$> 10^{13}\ \Omega \cdot$cm
Relative dielectric constant	9–9.2

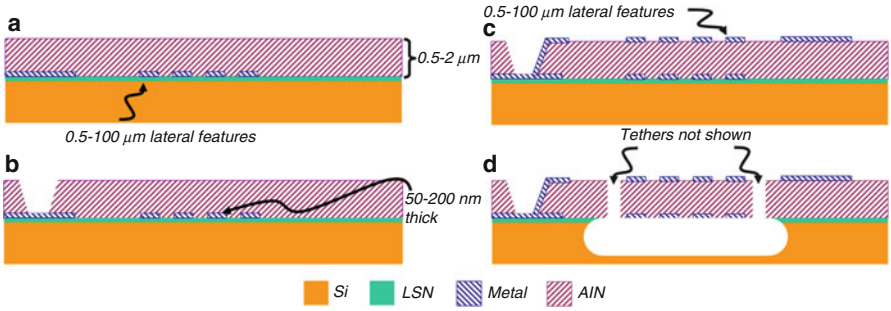

Fig. 2.9 Schematic representation of the fabrication process used to make AlN CMR resonators: (**a**) low stress nitride (LSN) is deposited on a silicon wafer and bottom metal electrodes are patterned on it; the AlN is sputter-deposited on the entire wafer. (**b**) Via access to the bottom electrode is etched into the AlN layer. (**c**) Top metal electrode and via interconnects are patterned; (**d**) AlN is dry etched and the structure is selectively released in XeF_2

2.1.4 AlN Piezoelectric Films and Microfabrication Process for Contour-Mode Resonators

One of the key elements that have made possible the demonstration of the piezoelectric CMR technology has been the repeatable deposition of thin and highly oriented AlN films on silicon substrates by means of sputtering deposition techniques. It is, in fact, thanks to the recent commercial deployment of AlN-based FBAR filters that deposition tools for piezoelectric AlN have become more readily available. All the experimental results presented in this chapter have been obtained by using either AMS or Tegal sputtering equipment. The quality of the AlN film is critical to the demonstration of the CMR devices. In particular, it is important to monitor not only the piezoelectric coefficient of these films but also their electrical resistivity, dielectric constant, and degree of crystallinity. The key properties that have been achieved in AlN films are summarized in Table 2.2.

The fabrication process for AlN CMR devices is relatively straight forward and uses 4–5 masks and conventional integrated circuits (IC) processing techniques. As shown in Fig. 2.9, the process consists in sputter-depositing highly oriented metal electrodes (generally aluminum or platinum, but other conductors can also be used) and patterning them into the desired shape. The metals can be deposited directly on a high-resistivity silicon substrate or a thin dielectric layer. In most cases, a thin adhesion layer, such as titanium, is used. The AlN film is then sputter-deposited,

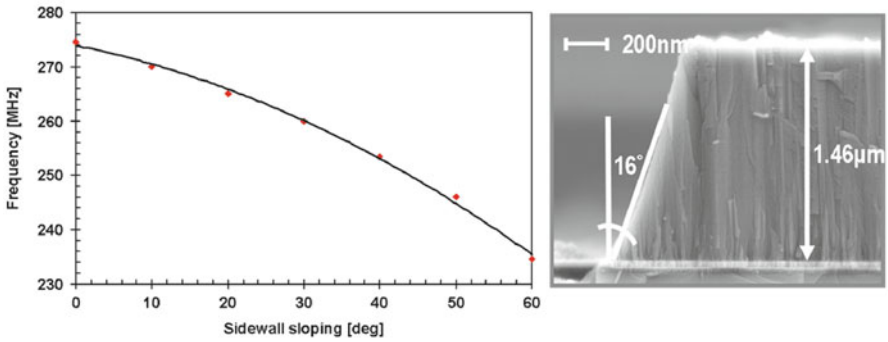

Fig. 2.10 ANSYS simulation of the effect of sidewall slope (measured from the vertical as in the SEM inset) on the frequency of a 275-MHz laterally vibrating CMR device. ©2006 IEEE. Reprinted, with permission, from [2]

and vias to make contact to the bottom layer are formed by wet etching of AlN by phosphoric acid (H_3PO_4) or other less aggressive solutions based on potassium hydroxide (KOH). The top metal electrode is then sputter-deposited and patterned by means of lift-off or dry etching. The AlN is etched by a Cl_2-based dry etch step to define the main resonator body (rectangular or annular body) and access the supporting silicon layer that needs to be removed for freeing the structure from the substrate. The release is performed by dry etching in XeF_2, which selectively removes silicon without attacking any of the other layers that form the resonant device. An electroplating step can be included (generally Au) to completely fill the via and thicken the lines used to route the signal to the resonator, therefore minimizing any ohmic loss modeled by R_S in Fig. 2.4.

The most critical steps that affect the reliable manufacturing of the CMR devices are those that directly set the resonance frequency of the structure. In fundamental mode devices, the etching of AlN and electrode definition is used to accurately prescribe the device center frequency. For most filtering applications, ±1,000 parts per million (ppm) accuracy is considered sufficient to attain high yield without the need of post-process trimming. Using 1,000 ppm as a reference mark, fundamental mode devices can be readily fabricated up to approximately 400 MHz using a 0.5-μm lithographic tool. At the same time, accurate and repeatable definition of the etching step needs to be guaranteed. In fact, as shown in Fig. 2.10, the center frequency of fundamental mode devices is highly affected by changes in the sidewall slope.

As already anticipated in Sect. 2.1, the use of higher order mode devices drastically relaxes the manufacturing constraints for higher frequency devices and renders the fabrication of devices up to approximately 3 GHz very reliable using a 360-nm line lithography tool. The use of the electrode pitch, rather than the overall plate dimension, to prescribe the frequency of operation of the device makes the CMR technology more robust against process variations. It is also easier to define smaller features in thinner electrode materials (50–200 nm thick) rather than thicker

(1–2 μm) AlN films. The use of *n* electrodes further reduces the device sensitivity to the AlN etching step, since variations in the plate profile are weighted over the number of electrodes.

The CMR sensitivity to variations in the electrode thickness is of the same order of magnitude of SAW and FBAR devices, but CMR exhibits drastically reduced sensitivity to changes in the AlN thickness over FBAR technology. In fact, the AlN thickness plays a second-order role in defining the center frequency of CMR devices.

2.2 Aluminum Nitride Contour-Mode MEMS Oscillators

One of the largest application areas for resonators is timing. Although few electronic components have stood the test of time better than quartz crystal oscillators, MEMS-based solutions have emerged as a very promising and competitive alternative due to their small form-factor, high operating frequency and especially the possibility to be fully integrated with ICs. A key advantage of the AlN CMR technology over other MEMS implementations is its ability to offer impedances that can be directly interfaced with standard CMOS oscillator configurations such as Pierce and Colpitts (Fig. 2.11). Other unique features of the AlN technology are the ability of attaining

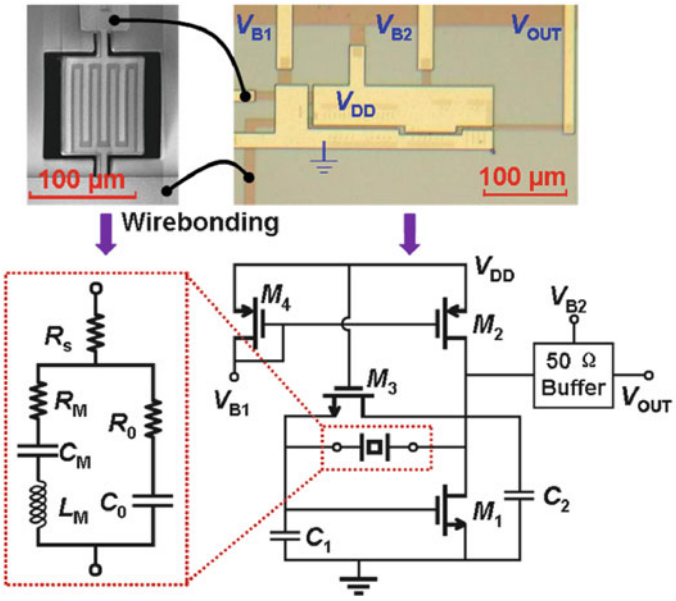

Fig. 2.11 Images of the AlN CMR device and oscillator circuit that were wirebonded. At the bottom, their equivalent electrical models are shown. ©2008 IEEE. Reprinted, with permission, from [13]

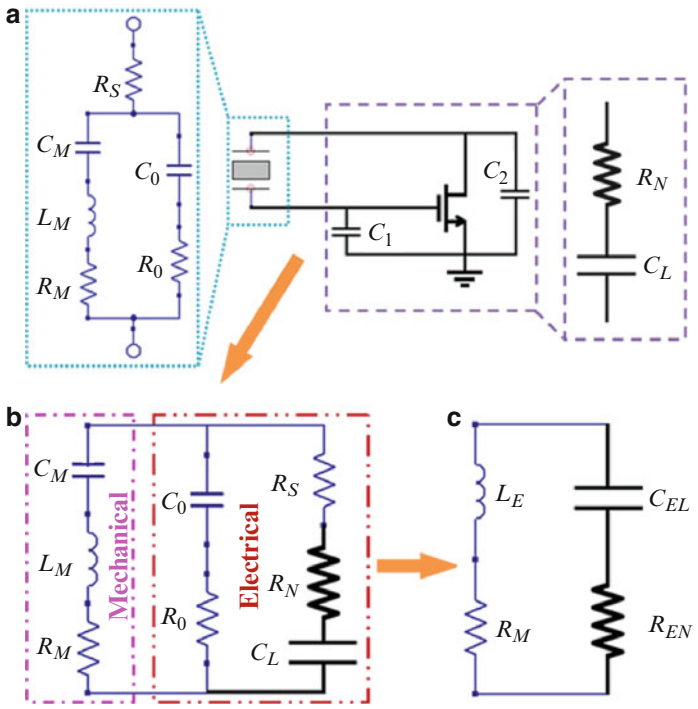

Fig. 2.12 Equivalent circuits of the Pierce oscillator operating near the oscillation frequency. (**b**) Separation of the motional (mechanical) and electrical parts of the circuit. (**c**) Rearrangement to simplify the oscillator analysis. ©2008 IEEE. Reprinted, with permission, from [13]

multiple frequencies of operation on a single chip and withstanding large power. In fact, the amplitude-frequency coefficient of an AlN resonator at 222 MHz was measured to be ~8 ppm/mA^2, which is comparable to quartz crystal devices and about eight orders of magnitude better than electrostatic devices operating at few MHz.

The Vittoz method [16] has been commonly used for the small-signal analysis of three-point oscillators with crystal resonators and can also be applied to the design of multifrequency CMR devices. However, since electrode and substrate parasitics (e.g., R_S and R_0 in Fig. 2.4) play a significant role in device performances at the microscale, the Vittoz method needs to be modified accordingly to take these parameters into account and design oscillators based on MEMS resonators. As shown in Fig. 2.12a, the active branch of the Pierce oscillator can be considered equivalent to a negative resistance R_N in series with a load capacitance C_L, the values of which are given as follows:

$$R_N = -\frac{g_m}{\omega^2 C_1 C_2}, \tag{2.12}$$

$$C_L = \frac{C_1 C_2}{C_1 + C_2}, \tag{2.13}$$

where $\omega = 2\pi f$ is the operating frequency, g_m is the small-signal transconductance of the NMOS transistor, and C_1 and C_2 are respectively the gate-source and drain-source capacitances. The Pierce circuit operating near the oscillation frequency f_o can be simplified to a four-component network that consists of an effective inductance L_E, an effective load capacitance C_{EL}, an effective negative resistance R_{EN}, and the motional resistance, R_M, of the resonator, as shown in Fig. 2.12c. These effective values can be derived as

$$L_E \approx \frac{2(\omega - \omega_s)}{\omega^2 \omega_s C_M} = \frac{2p}{\omega^2 C_M} \qquad p = \frac{\omega - \omega_s}{\omega_s} \ll 1, \tag{2.14}$$

$$Z_{ET} = [R_N + R_S + 1/(j\,\omega C_L)] \,||\, [R_0 + 1/j\,\omega C_0)], \tag{2.15}$$

$$R_{EN} = \text{Real}\{Z_{ET}\} \quad \text{and} \quad C_{EL} = -1/[\omega \cdot \text{Imag}\{Z_{ET}\}], \tag{2.16}$$

where p is defined to be the frequency pulling factor, which indicates the relative amount of frequency pulling above the series resonant frequency, ω_s, of the resonator, Z_{ET} is the total impedance of the electrical part in Fig. 2.12b, "||" means parallel electrical connection, and "Real" and "Imag" denote real and imaginary parts of a complex number, respectively. By equating R_{EN} with $-R_M$ and solving (2.16), the critical transconductance, g_{mc}, and operating frequency, ω_c, for starting oscillations can be estimated.

Using these analytical models for the resonator and circuit, it is possible to highlight how the key resonator components (Q, k_t^2, frequency, geometry, and parasitics) affect power consumption in the oscillator circuit.

In Fig. 2.13, the critical transconductance, g_{mc}, is plotted as a function of Q, k_t^2, nL, T, f, and R_0, while all other parameters are kept fixed. The primary source of power consumption in the oscillator is due to the motional resistance, R_M, of the resonator. As already stated, R_M is inversely proportional to $k_t^2 \cdot Q$, a figure of merit of the resonator.

Figure 2.12b shows the dependence of g_{mc} on the resonator geometry. Since the sub-resonator width, W, is used to define the resonant frequency in the piezoelectric AlN CMR technology, only the effective length, nL, and thickness, T, are available for design optimization. R_M is inversely proportional to nL, and this explains why the value of g_{mc} increases with a decrease in nL. On the other hand, if nL is large, then the impedance of C_0 is low and R_0 becomes a significant part of power loss in the circuit. Therefore, parasitics need to be taken into account since they can significantly affect the oscillator power consumption. The resonator thickness, T, has little effect within the optimal nL region, and therefore, any thickness could be employed. To ensure good material quality of the sputter-deposited AlN and low device impedance $(1/j\,\omega C_0)$, $T = 2\,\mu$m, $L = 200\,\mu$m, and $n = 3$ were first employed for the demonstration of a 222-MHz oscillator.

In order to demonstrate a single circuit (amplifier) design that works for multi-frequency resonators, it is important to analyze how g_{mc} varies as a function of operation frequency, f, and parasitic resistance R_0 (Fig. 2.13c). As (2.12) implies, given a certain R_N, g_{mc} rapidly increases with frequency. At high frequencies,

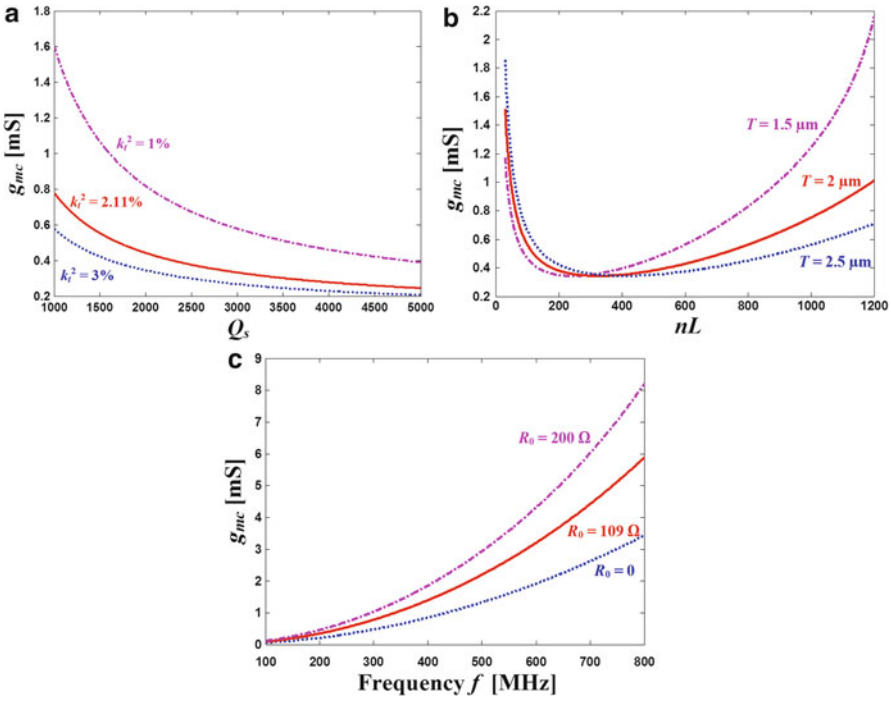

Fig. 2.13 (**a**) g_{mc} vs Q_s for different k_t^2; (**b**) g_{mc} vs nL for different T; (**c**) g_{mc} vs f for different R_0. ©2008 IEEE. Reprinted, with permission, from [13]

the effect of R_0 becomes more pronounced as already illustrated in Fig. 2.12b (the impedance associated with C_0 lowers with increasing frequency). Overall, the dominating parameter that sets the single Pierce circuit design for multifrequency oscillators is the operation frequency.

In the first AlN CMR-based multifrequency oscillator prototype, a maximum frequency of 500 MHz was used and set the maximum power consumption to 10 mW in the AMIS 0.5 µm 5 V CMOS process. AlN contour-mode resonators at 176, 222, 307, and 482 MHz were connected to the same Pierce oscillator circuit (Fig. 2.11). These devices attained phase noise values between −88 and −68 dBc/Hz at 1 kHz offset frequency from the carriers and phase noise floors as low as −160-dBc/Hz at 1 MHz offset. The characteristic phase noise response of a 222-MHz resonator is shown in Fig. 2.14. The same phase noise of the 222-MHz oscillator was also integrated from 12 kHz to 20 MHz, and the RMS jitter found to be 74 fs. This value is comparable to what is available from SAW-based oscillators working in the same frequency range.

As shown in Fig. 2.14, the phase noise performance of this first oscillator prototype is limited by the circuit layout, and further optimization is possible. In addition, by utilizing a more modern CMOS technology, lower power consumption will also be possible.

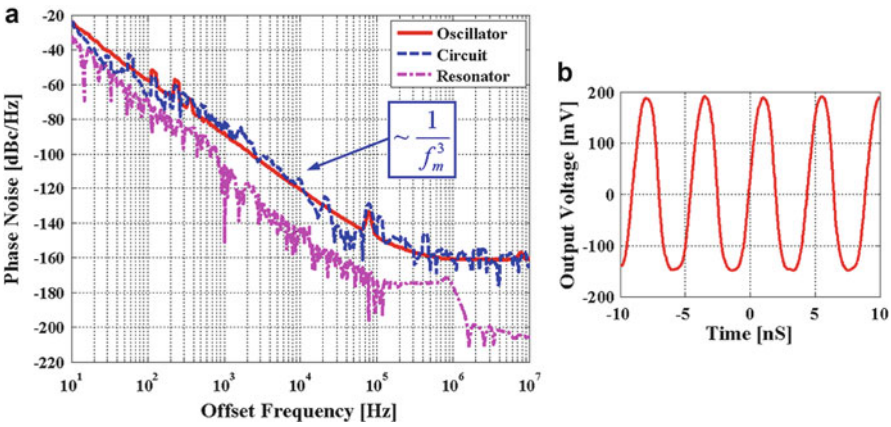

Fig. 2.14 (**a**) Closed-loop phase noise plots of the 222-MHz oscillator (*solid line*) and the corresponding contributions from circuit without resonator (*dashed line*) and resonator (*lower dash-dotted line*). (**b**) Steady-state oscillator output voltage showing waveform asymmetry. Voltage clipping in the oscillator core circuit is likely to be a major contributor to the close-to-carrier phase noise by up-converting baseband flicker noise into the carrier signal band. ©2008 IEEE. Reprinted, with permission, from [13]

2.3 Aluminum Nitride Contour-Mode MEMS Filters

Given their high frequency of operation and ability to achieve high Q and low impedance in a small form factor, the CMR technology is extremely suited for the definition of RF MEMS filters.

Approximately 10–15 filters are present in most cell phones today, and these numbers are steadily growing due to the need for multiple bands and applications. SAWs and FBARs are main stream in the cell phone bands but are limited in their ability to provide solely few closely spaced frequencies on the same chip. The AlN CMR technology can instead span a broader range of frequencies on the same die (MHz to GHz), therefore providing a possible answer to the need of compact multiband and multifrequency components that currently limit the wireless communication industry.

Filters are made out of an array of resonators coupled either electrically or mechanically. CMR resonators are employed to precisely locate poles and zeros and achieve flat pass band, sharp roll off, narrow bandwidth and low loss filters. The next sections describe different methods to synthesize MEMS filters out of CMR devices.

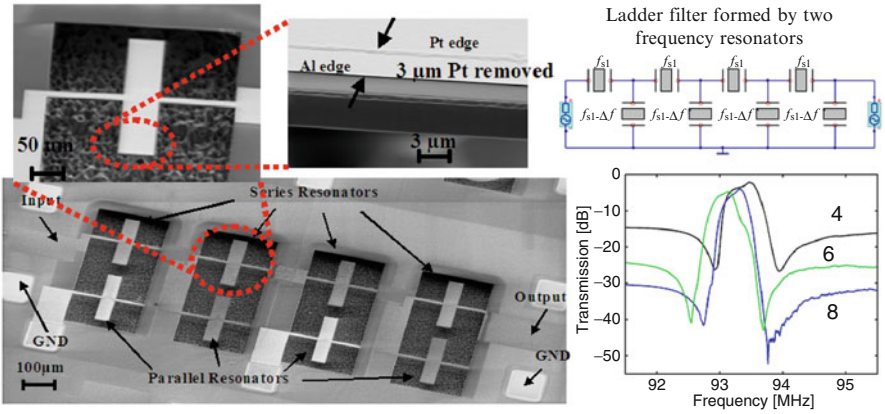

Fig. 2.15 SEM of an eight resonator ladder filter using CMR devices. The insets show the circuit schematic of the filters (which highlights that two resonators with different frequencies are required to implement such filter) and the filter experimental performance when using a different number of resonators. ©2007 IEEE. Reprinted, with permission, from [17]

2.3.1 Electrically Coupled AlN Contour-Mode Filters

Simpler to implement than any other coupling technique, electrical coupling requires routing of electrical signals from one CMR device to the next without the need for any external electrical components. It is a robust coupling technique commonly used with MEMS devices and more generally with mechanical filters. AlN CMR devices have been arranged in conventional ladder topologies to demonstrate filter prototypes around 94 MHz [17] (Fig. 2.15).

Another very effective filter implementation using electrically cascaded AlN CMR devices has been shown to work at 94 and 271 MHz [18] (Fig. 2.16). In this design, three or four two-port AlN resonators are connected in series and coupled by their intrinsic capacitance to realize high-order filtering. This solution offers the possibility to design filters with good shape factors and off-band rejection without the need for different frequency devices as it is instead deemed necessary in the ladder configuration of Fig. 2.15. This solution reduces the complexity of the filter design and fabrication and ultimately improves the device yield. Also, this coupling technique intrinsically provides for narrow bandwidths (0.1–0.4%), and makes the AlN technology a very good fit for the demonstration of channel-select filters. As will be highlighted in the next section, channel selection, instead of band selection, represents a significant paradigm shift in the way current RF front ends are designed and can drastically reduce power consumption and architecture complexity.

This filter topology can be synthesized by cascading same frequency devices. The use of two port devices (Fig. 2.6) eliminates the need for external coupling capacitors, since the intrinsic device capacitance to ground can be employed. For example, in a third order filter, three fundamental resonant modes (3 poles) can be

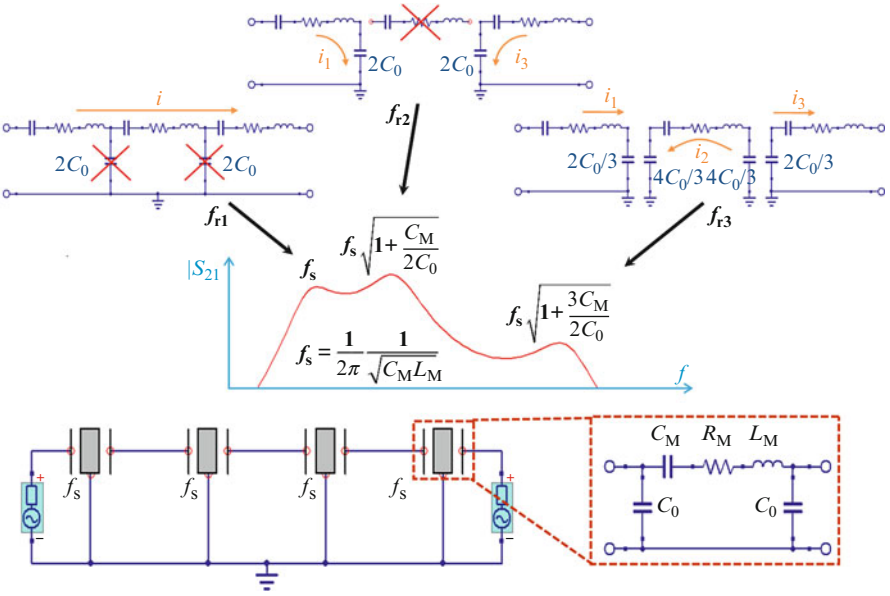

Fig. 2.16 Schematic representation and principle of operation of an electrically coupled filter formed by same frequency CMR devices cascaded one after the other. The intrinsic capacitance to ground of each two-port resonator provides for the coupling element required to synthesize the filter pass band

synthesized by cascading three two-port devices (Fig. 2.15). The filter bandwidth is approximately set by the spacing between the first and third mode of vibration and is given by

$$FBW_{3\text{dB}} \approx \frac{f_{r3} - f_{r1}}{f_{r1}} = \sqrt{\frac{3C_M}{2C_0} + 1} - 1 \approx \frac{3}{\pi^2} k_t^2, \tag{2.17}$$

where f_{r3} and f_{r1} are defined as in Fig. 2.16, and C_M and C_O are the equivalent parameters of the two-port resonator. It is important to note that the k_t^2 of the device sets the filter bandwidth.

To maximize power transmission from input to output, the filter needs to be terminated by a proper resistance R_T, which is primarily determined by the input impedance of the filter and can be approximated by the impedance magnitude of the input transducer, C_0, of the first resonator stage. It can in fact be proven that in order to attain maximum power transfer (i.e., minimum insertion loss), R_T needs to be set approximately equal to:

$$R_T \approx \left| \frac{1}{j\,\omega_s C_0} \right|. \tag{2.18}$$

Approximately, an nth-order filter at resonance is equivalent to a circuit of n series resistors, each having value of R_M, with two additional intrinsic capacitors C_0 at the two ends. It can therefore be shown that the filter insertion loss IL can be expressed as

$$IL(\text{dB}) \approx -20\log_{10}\left(\frac{4}{4+n\pi^2/k_t^2 Q}\right). \tag{2.19}$$

Equation (2.19) shows how filter IL exclusively depends on $k_t^2 \cdot Q$ and further emphasize the importance of this figure of merit. When designing a filter, another important parameter to take into account is its out-of-band rejection. In this filter topology, the resonators behave as open circuits and the overall equivalent circuit corresponds to the two input capacitors to ground and a feedthrough capacitance, C_F. Then, the out-of-band rejection is determined by C_F, which is generally set by substrate parasitics and can be controlled by proper routing and spacing of the input and output electrodes. In this filter, the out-of-band rejection is given by:

$$Rej.(\text{dB}) \approx -20\log_{10}\left(\frac{C_f}{C_0}\right) - IL. \tag{2.20}$$

C_F can generally be limited to values around few fF, therefore providing for large out-of-band rejection in excess of 60 dB (Fig. 2.17).

The filter shape factor is defined as

$$X\text{dBShape Factor} = \frac{\text{Filter bandwidth at XdB below IL}}{\text{Filter bandwidth at 3dB below IL}}, \tag{2.21}$$

where X is the number of dB lower than the IL at which the filter bandwidth is measured. The shape factor primarily depends on the number of poles forming the filter (the order of the filter) and to a second degree on the Q of the device.

Figure 2.17 shows an experimental demonstration of the described filter topology, for which a bandwidth of 0.34%, IL of 4.2 dB, and rejection >60 dB were attained at 271 MHz.

2.3.2 *Mechanically Coupled AlN Contour-Mode Filters*

Mechanical coupling techniques require the use of mechanical links between different resonators to synthesize band-pass filters. Extensional beams or solid quarter wave couplers (typical of monolithic filters) have been employed to realize these types of filter. The principle of operation can intuitively be understood by looking at the equivalent lumped mechanical model of a resonator. As shown in Fig. 2.18, two resonators coupled by a mechanical element (generally represented by a spring and a mass) behave as a coupled system of oscillators and are

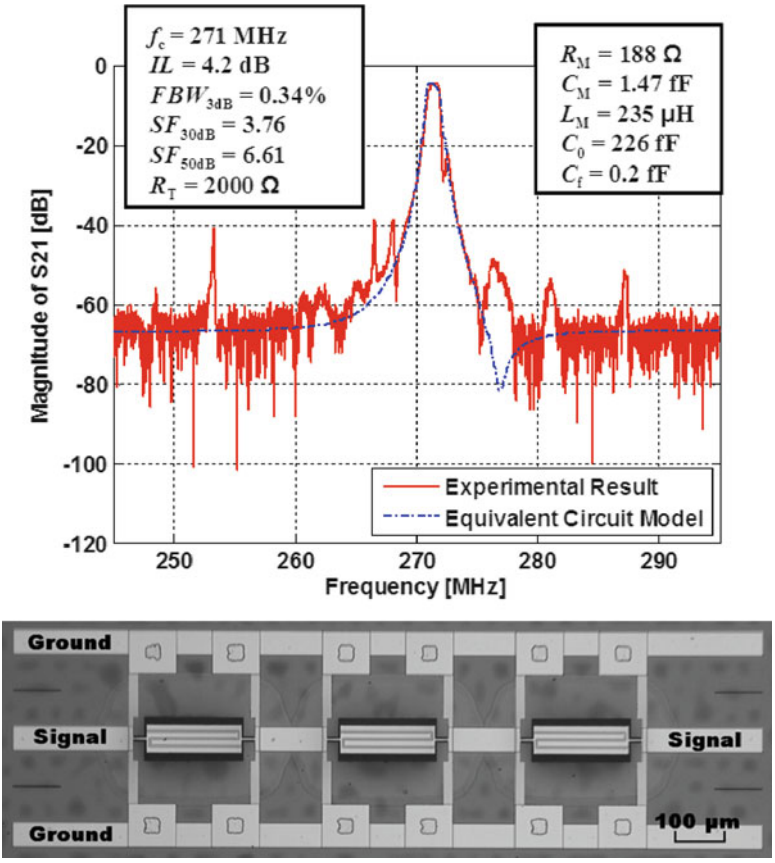

Fig. 2.17 Experimental response of a 271-MHz filter formed by electrically cascaded two port AlN CMR devices. Inset shows an image of the same filter. ©2007 IEEE. Reprinted, with permission, from [14]

therefore characterized by two possible modes of oscillation. By properly defining the stiffness (and mass) of the coupling element, the frequency spacing (i.e., the bandwidth of the filter) between the modes of vibration can be appropriately set. Once this is understood, then standard techniques used to describe *IL*, rejection, and shape factor of electrically coupled filters can be applied to the mechanically coupled version. It is in general hard to control the location of different poles and the elimination of spurious modes in mechanically coupled filters of order greater than two. It is because of this reason that the most promising mechanical filters are limited to the second order.

A specific case of CMR mechanically coupled filter is the one shown in Fig. 2.19, for which a set of two width-extensional rectangular plates are coupled via a length-extensional beam. The two mechanical resonators are modeled in the same

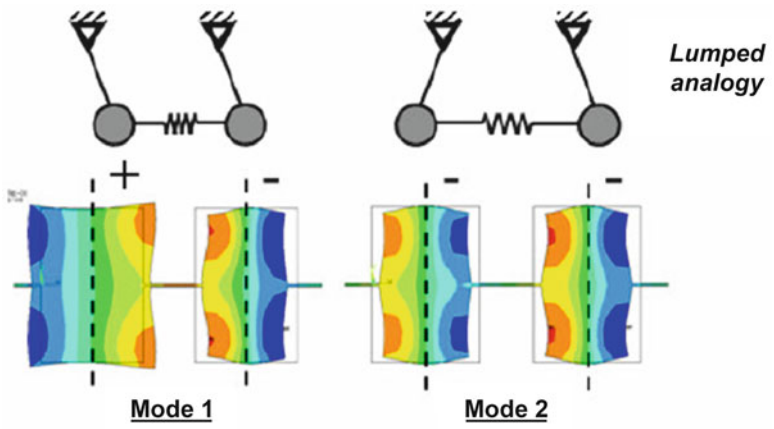

Fig. 2.18 Schematic representation of a second order coupled system of oscillators. *Top* represents an equivalent lumped analogy, whereas the *bottom* represents the distributed version formed by CMR devices. ©2006 IEEE. Reprinted, with permission, from [19]

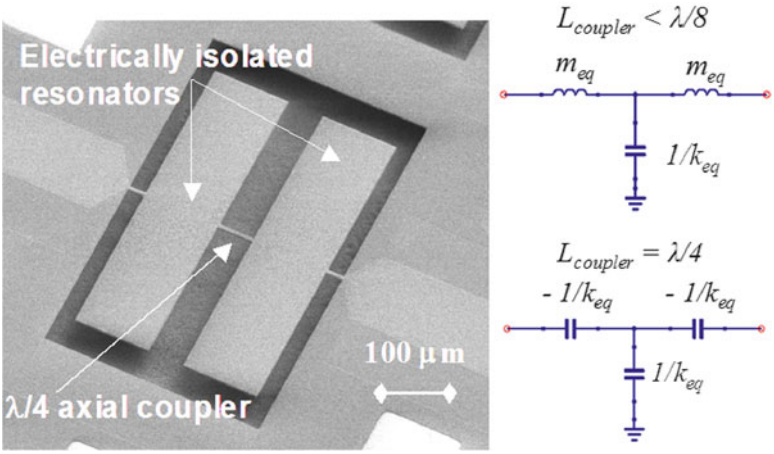

Fig. 2.19 SEM of a mechanically coupled CMR filters formed by two width-extensional resonators coupled by a $\lambda/4$ axial bar. The equivalent T network representation of the coupling elements is shown for different length of the coupler in the right part of the figure. ©2006 IEEE. Reprinted, with permission, from [19]

way shown in Fig. 2.1. The coupling element is instead treated as a mechanical waveguide and is described by the following relationship:

$$\begin{bmatrix} v_1 \\ F_1 \end{bmatrix} = \begin{bmatrix} \cos\left(kL_{coupler}\right) & \frac{j\sin\left(kL_{coupler}\right)}{Z_{AC}} \\ jZ_{AC}\sin\left(kL_{coupler}\right) & \cos\left(kL_{coupler}\right) \end{bmatrix} \cdot \begin{bmatrix} v_2 \\ F_2 \end{bmatrix}. \tag{2.22}$$

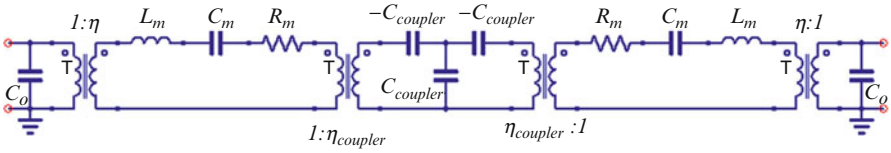

Fig. 2.20 Equivalent electrical model of a second order mechanically coupled filter

This matrix form relates the input and output mechanical vectors representing force, F, and velocity, v (subscripts 1 and 2 refers to the two ends of the bar). In (2.22), k is the wave vector, Z_{AC} is the acoustic impedance of the bar ($Z_{AC} = \sqrt{E\rho} \cdot A$, where E and ρ are the Young's modulus and density of the material that forms the coupling bar and A is the cross-sectional area of the bar), and $L_{coupler}$ is the length of the coupling element. Depending on the length of the coupler and the frequency of operation, the coupling bar can be modeled by a T network formed by combinations of equivalent masses and springs (as in Fig. 2.19). Generally, in order to minimize manufacturing dependence of the bandwidth of the filter to the dimensions of the coupler, the coupler length is chosen to be a quarter wavelength of the acoustic frequency of operation of the resonator. In this manner, the T network is formed by three springs of the same value but different signs. The spring is a representation of the equivalent stiffness of the coupling bar. As in Sect. 2.1, a mechanical stiffness corresponds to a capacitor in an electrical circuit representation.

Another element, a mechanical transformer, needs to be included between the resonator and the coupling element (Fig. 2.20), in order to take into account any difference in velocity between the coupling point and the point at which the equivalent lumped model of the resonator has been calculated. In this specific case, the resonator is coupled at the point of maximum velocity, and therefore, the transformer has a 1:1 turn ratio (it would otherwise be the ratio of the velocities at the two aforementioned points).

According to the equivalent model described in Fig. 2.20, it is possible to derive an expression of the fractional bandwidth of the filter, which depends exclusively on the spring constant of the resonator and the one of the coupling element:

$$FBW_{3dB} = \sqrt{1 + \frac{C_M}{C_{coupler}}} - \sqrt{1 - \frac{C_M}{C_{coupler}}} \approx \frac{C_M}{C_{coupler}} \approx \frac{k_{coupler}}{k_{eq}}. \tag{2.23}$$

It is important to note that the fractional bandwidth of the filter can be set lithographically by varying the size of the coupling bar. This is a distinct feature of mechanically coupled filters over purely electrically coupled ones. Because of limitations on ripple and IL, also the bandwidth of mechanically coupled filters can hardly be extended beyond the value of k_t^2.

Termination resistance and insertion loss can be computed in a manner similar to the one presented for electrically coupled filters. The rejection is generally

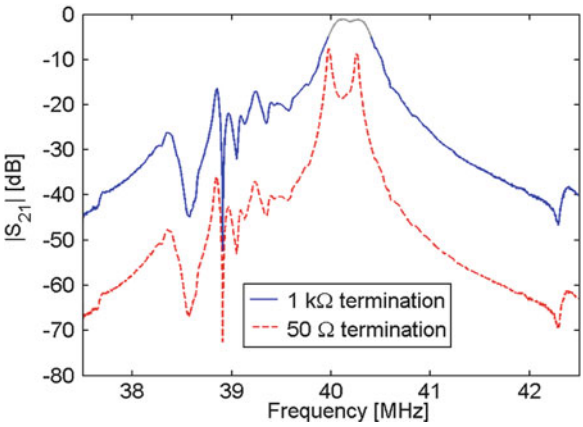

Fig. 2.21 Experimental response (S_{21}) of a 40-MHz mechanically coupled filter. The unterminated response highlights the two peaks formed by the coupled system of oscillators. ©2006 IEEE. Reprinted, with permission, from [19]

determined as in (2.20), although, depending on the strength of the mechanical coupling, it can be limited by the mechanical current induced by the out-of-resonance displacement of the coupled resonator system.

The typical response of a second-order mechanically coupled filter formed by AlN CMR width-extensional devices is shown in Fig. 2.21, for which *IL* of 1.5 dB at 40 MHz have been attained.

2.4 Applications for Aluminum Nitride Contour-Mode MEMS Resonator Technology

A unique advantage of the AlN CMR technology over other resonant devices is the ability to provide for multiple frequencies of operation on the same silicon chip. If integrated with switches (either micromechanical or electronic), the CMR technology can be employed for the demonstration of highly reconfigurable monolithic RF front ends that will represent a paradigm shift in the way RF signal processing is performed. By taking advantage of the possibility to array banks of switched narrowband filters at different frequencies, previously envisioned, but never implemented, RF front ends can be made possible. For example, channel selection in wideband code-division multiple-access (W-CDMA) can be implemented by arraying twelve distinct filters with a fractional bandwidth of approximately 0.25% (twelve 5-MHz channels spaced 200 kHz apart around 2 GHz). The filter bank can be employed to directly select the carrier signal before downconversion in low IF or direct conversion receivers (Fig. 2.22). The proposed architecture significantly reduces the power consumption in low noise amplifiers (LNAs), mixers, and voltage-controlled-oscillators (VCOs) by improving rejection directly at the channel and relaxing the linearity requirements on the electronic components and the phase noise of the oscillator.

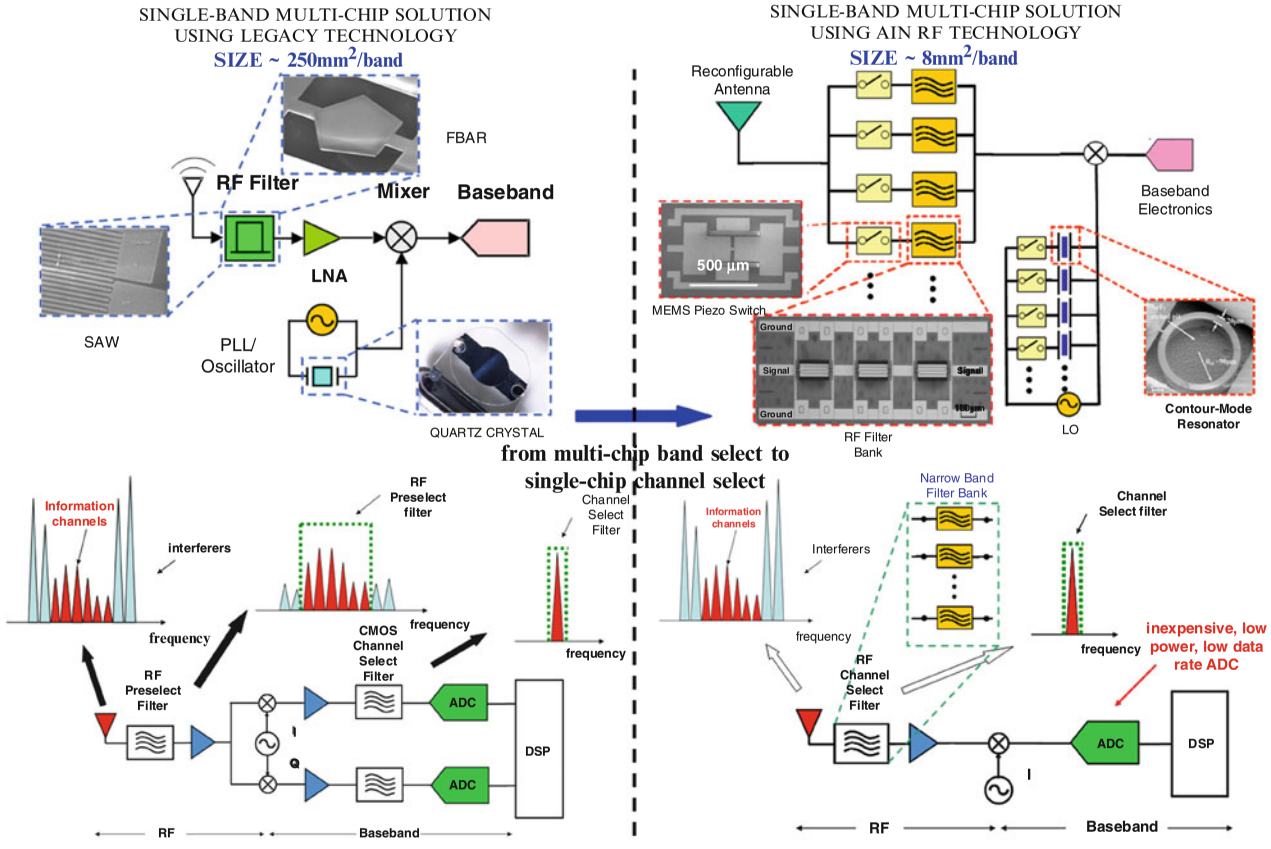

Fig. 2.22 Comparison between legacy technology (SAW, FBAR, and quartz crystal based) RF front-end and CMR enabled next-generation radio architectures. Reprinted with permission from [8]. ©2009, American Vacuum Society

Future software-defined radios (SDR) will also have similar requirements in terms of front-end filtering. In order to reduce power consumption in the ADC, direct channel sampling at the antenna location needs to be implemented by means of switched arrays of narrowband filters (Fig. 2.22). At the same time, broadband and direct frequency synthesis via high Q oscillators can be implemented with the same CMR technology. A very large area of the available spectrum (e.g., 300 MHz–5 GHz) needs to be covered: large scale integration (LSI) of microelectromechanical CMR devices will make this possible.

2.5 Concluding Remarks and Future Directions

This chapter introduced an emerging MEMS resonator technology that has been just recently developed in a university setting. The AlN CMR technology bears great promises of constituting one of the solutions for next-generation communication devices. Novel RF architectures that take advantage of LSI of AlN CMR MEMS devices have been presented as revolutionary designs that will significantly reduce power consumption in modern communication systems and will enable next-generation software-defined radios. Despite the noteworthy progress that has been made with the AlN CMR MEMS technology, there are still significant challenges that need to be addressed in terms of characterizing these devices to meet specifications for power handling, temperature stability and packaging, as well as improving the resonator quality factor and electromechanical coupling. Furthermore, more recent demonstrations of CMR devices at microwave frequencies (5–10 GHz) [20] open an entire new range of applications for this technology. These topics are the subject of ongoing research.

Acknowledgements The author would like to thank his students Chengjie Zuo, Nipun Sinha, Carlos Perez, Matteo Rinaldi, and Chiara Zuniga without whose contribution many of the results presented in this chapter would have not been possible. The author would also like to extend his greatest gratitude to Dr. Philip J. Stephanou, Dr. Justin P. Black, and Prof. Albert P. Pisano, who all share the same passion for this MEMS resonator technology and first made possible its demonstration.

References

1. R. Ruby, P. Bradley, I. J. Larson, Y. Oshmyansky, D. Figueredo, in *2001 IEEE International Solid-State Circuits Conference* (IEEE, 2001), 2001 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC (Cat. No.01CH37177), pp. 120–121
2. G. Piazza, P.J. Stephanou, A.P. Pisano, J. Microelectromech. Syst. **15**(6), 1406 (2006)
3. C.T.C. Nguyen, IEEE Trans. Ultrason. Ferroelectrics Freq. Contr. **54**(2), 251 (2007)
4. S. Humad, R. Abdolvand, G.K. Ho, G. Piazza, F. Ayazi, in *IEEE International Electron Devices Meeting 2003*, IEEE, 2003, IEEE International Electron Devices Meeting 2003, pp. 39–3
5. D. Weinstein, S.A. Bhave, in *2007 International Electron Device Meeting*, 2007, pp. 415–418

6. K.M. Lakin, G.R. Kline, R.S. Ketcham, S.G. Burns, in *IEEE 1986 Ultrasonics Symposium Proceedings*, IEEE, Williamsburg, VA, USA, 1986, IEEE 1986 Ultrasonics Symposium Proceedings (Cat. No.86CH2375-4), pp. 371–375
7. M.A. Dubois, C. Billard, J.F. Carpentier, P. Vincent, G. Parat, C. Muller, in *2005 IEEE Ultrasonics Symposium*, *Proceedings - IEEE Ultrasonics Symposium*, vol. 1 (Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ 08855-1331, United States, Rotterdam, Netherlands, 2005), *Proceedings - IEEE Ultrasonics Symposium*, vol. 1, pp. 85–88
8. G. Piazza, J. Vac. Sci. Tech. A **27**(4), 776 (2009)
9. G. Piazza, P.J. Stephanou, A.P. Pisano, Solid State Electron. **51**(11-12), 1596 (2007)
10. R.A. Johnson, *Mechanical Filters in Electronics* (Wiley, New York, 1983)
11. H.A.C. Tilmans, J. Micromech. Microeng. **6**, 157 (1996)
12. H.A.C. Tilmans, J. Micromechanics Microengineering **7**(4), 285 (1997)
13. C. Zuo, N. Sinha, J. Van der Spiegel, G. Piazza, in *Frequency Control Symposium, 2008 IEEE International*, 2008, pp. 402 –407. DOI 10.1109/FREQ.2008.4623028
14. C. Zuo, N. Sinha, M. Pisani, C. Perez, R. Mahameed, G. Piazza, in *IEEE 2007 Ultrasonics Symposium,* 2007, pp. 1156–1159. DOI 10.1109/ULTSYM.2007.292
15. G. Piazza, A.P. Pisano, Sensor Actuator A Phys. **136**(2), 638 (2007)
16. E.A. Vittoz, M.G.R. Degrauwe, S. Bitz, IEEE J. Solid State Circ. **23**, 774 (1988)
17. G. Piazza, P.J. Stephanou, A.P. Pisano, J. Microelectromech. Syst. **16**(2), 319 (2007)
18. C. Zuo, N. Sinha, M.B. Pisani, C.R. Perez, R. Mahameed, G. Piazza, in *IEEE 2007 Ultrasonics Symposium,* 2007, pp. 1156–1159
19. P. Stephanou, G. Piazza, C. White, M. Wijesundara, A. Pisano, in *Micro Electro Mechanical Systems, 2006. MEMS 2006 Istanbul. 19th IEEE International Conference on* (2006), pp. 906–909. DOI 10.1109/MEMSYS.2006.1627947
20. M. Rinaldi, C. Zuniga, G. Piazza, in *22nd IEEE International Conference on MicroElectroMechanical Systems, IEEE MEMS 2009*, 2009, pp. 916–919

Chapter 3
Nanoelectromechanical Systems (NEMS)

Adrian Ionescu

Abstract This chapter focuses on NEM devices and their technologies and applications for information processing with special emphasis on NEM switches and resonators. Top-down and bottom-up fabrication of NEMS are presented. Then NEM switches for ultralow-standby-power digital applications and their fabrication and design challenges are described; silicon or metal nanowires and carbon nanotube switches are able to achieve interesting figures of merit for nanoscale electromechanical information processing (logic and memory functions). The chapter continues with NEM resonators; the extreme scaling requires active detection such as resonant gate/body transistors. Finally, the concept of NEMS-based radio is introduced and its promises and challenges are discussed.

3.1 Electromechanical Information Processing

Today, the engineering of Nanoelectromechanical systems (NEMS) for analog, radio frequency (RF) and sensing applications became a reality due to the tremendous progress and maturity in microelectromechanical systems (MEMS) fabrication and, especially, because of the convergence of MEMS and silicon CMOS processes. Of particular importance for the progress of NEMS are the last decade achievements in the domain of radio-frequency MEMS (RF MEMS) that opened many opportunities for improved circuit performance at high frequency (MHz to 100 GHz) combined with low-power static consumption and affordable technology costs [1, 2]. The aggressive miniaturization of MEMS was possible thanks to silicon technology that, after year 2000, entered into the nanoelectronics era, featuring sub-100-nm device dimensions. Furthermore, the advancement of bottom-up technologies (especially nanowires and nanotubes) has created new opportunities for experimentally testing

A. Ionescu (✉)
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
e-mail: adrian.ionescu@epfl.ch

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_3,

novel NEMS principles, investigating and exploiting specific electromechanical properties at nanoscale beyond the simple device miniaturization, finally aiming at disruptive information processing and sensing applications.

In general, the mechanical information processing can be understood in two fundamentally different forms: (a) as a multistate logic, with the logic states dictated by a spatial configuration of movable objects, and (b) as vibrational modes of mechanical elements, based upon waves [3]. Devices with nanometer small size, like NEMS, generate supplementary interest for collective information processing and their use for building devices and systems that can address some of the main limitation of dense digital computation with CMOS such as standby power. Alternatively, the potential of NEMS for analog signal processing and sensing is enormous and difficult to completely imagine. Resonant nanostructures could embed today full equivalent circuit functions such as filtering or frequency references with figures of merit unreachable by MEMS or solid-state devices. Prospects for all-NEMS RF front-ends or even all-NEMS radios are transformed in experimental realities. The extremely small mass of NEM devices made possible the increase of their resonance frequencies close or beyond GHz and their mass sensitivities below the atogram. On the other hand, their small sizes raise great challenges in terms of manipulating and exploiting relatively poor signal-to-noise ratio and sensitivity to dynamic environment. This generated novel research aimed at new detection schemes (different from the traditional MEMS capacitive detection) and interfacing circuitry that can exploit the particular NEMS features. Nonlinearities and operating power levels are today intensely researched fields for NEMS applications.

While there are clear benefits of scaling down MEMS into NEMS, we should also fairly note that size reduction at nanoscale may not be a panacea for all applications [4]. Finally, the scaling limits of NEMS are at atomic and molecular scale, with electrical features set by switching or vibrational properties of molecules. Therefore, it is expected that research on NEMS and molecular electronics will have to meet in the future.

This chapter focuses on applications for information processing with special emphasis on NEM switches and resonators. It gives an overview on NEMS technology, NEM switches including silicon or metal nanowires and carbon nanotube switches, NEM resonators and detection schemes, and finally the concept of NEMS-based radio. Beyond the intrinsic NEM device performance, the success of NEMS-CMOS industrial products is conditioned by solving the still open issues of reliability and packaging. The heterogeneous or monolithic integration of NEMS and CMOS and hybrid NEMS-CMOS codesign are expected to trigger further functionality and performance at system level.

3.2 NEMS Technology: Top-Down or Bottom-Up?

NEMS technology requires the fabrication of nano-objects combined with surface micromachining [5] that involves the possibility of suspending and anchoring parts of the fabricated nanostructures based on deposited thin films. In general, a

Table 3.1 Various top-down fabrication techniques, their achievable minimum feature size and main advantages/issues

Top-down fabrication techniques	Minimum feature size	Advantages and issues
(DUV) optical lithography	∼30 nm	Batch processing, robust, controllable, complex
E-beam lithography [7]	∼5 nm	Low throughput, high resolution
Nanoimprinting [8]	∼10 nm	High throughput, low cost, alignment issues
Nanostencil [9]	∼50 nm	High throughput, low cost, alignment issues (∼1 μm)
Spacer or sidewall [10]	∼20 nm	High throughput, some limitation in design
Scanning probe lithography [11]	∼5 nm	Fast prototyping of individual devices with very small size
Reactive ion etching [12]	∼50 nm	Free of wet etching restrictions, suitable for mass production, profile and uniformity control needed at nanoscale
Metal deposition by liftoff [13]	∼10 nm	Removing metals hard to etch, surface microfabrication
Focused ion beam (FIB) [14]	∼30 nm	Versatile, fast, ion contamination issues

free-standing structure is obtained by using a sacrificial layer that is selectively etched. The fabrication techniques capable of making nanostructures can be grouped into three main categories [6]: (a) *top-down*, (b) *bottom-up*, and (c) *hybrid bottom-up/top-down*.

Top-Down processing is based on the optical lithography and etching batch-type processing of semiconductor industry, which gives high volumes and high productivity on silicon wafers. By using such an approach, NEMS can take benefits from all the maturity of CMOS processing and high-resolution DUV lithography, which is now industrially available at the 45-nm technology node. However, many NEMS developments are in a research phase and cannot afford such very complex and expensive advanced processing. Some of the most frequent structures used to build top-down NEMS are silicon and metal nanowires. A series of top-down fabrication techniques listed in Table 3.1 together with their minimum feature size, advantages, and issues are available: deep-ultraviolet (DUV) optical lithography, e-beam lithography [7], nanoimprinting [8], nanostenciling [9], spacer or sidewall techniques [10], scanning probe lithography (using a sharp tip in proximity to a sample to pattern nanometer-scale features) [11], reactive ion etching [12], metal deposition by liftoff [13], and focused ion beam [14]. Usually, these techniques are not singularly used to define a complete NEMS structure; a combination of two of more of the cited processes offers the best compromise for structures that combine different materials and different sizes of the active (movable or vibrating) regions

of the NEMS and their passive regions (contacts and interconnects, anchoring, packaging, etc.). The main characteristics of the top-down technique is that the minimum feature size of NEMS depends on the resolution of one of the before-mentioned processes. Finally, one can distinguish between direct and indirect top-down fabrication [7]. A typical indirect type of fabrication is the so-called sidewall technique [8]. In sidewall techniques, delineation of a large line pattern is made first by lithography and the sidewall of the line structure is conformally deposited with another material. By dissolving the line structure, the film deposited on its sidewall is left as a stand-alone feature, which can be of much smaller width, dictated only by the film thickness, than the original.

In the *bottom-up* approach, the key nanostructures used to fabricate NEMS are processed by a nonlithographical method: growth or self-assembly processing that can provide feature size much smaller than the resolution offered by the lithography. Alternatively, at comparable size, the bottom-up fabrication can offer much simpler and less expensive processes with a very high turnover [15]. Silicon nanowires (SiNWs) and carbon nanotubes (CNTs) are typical nanostructures than can be realized by bottom-up techniques with cross sections smaller than 50 nm. SiNWs seem particularly attractive since they can, in principle, be incorporated into existing processing flows and can be grown epitaxially at predefined locations. Growing SiNWs by the Vapor–Liquid–Solid (VLS) technique [15] has been found to be an attractive approach since VLS growth allows obtaining NWs with extremely small diameters. Moreover, there is no need for lithography and etching, which potentially leads to surface-roughness degradation. Other advantages are related to the fact that VLS-grown SiNWs can be doped in situ during growth, without the need of implantation and high-temperature anneals for dopant activation. However, issues concerning SiNWs controllability, variability, forming contacts and interfaces with desired electronic properties, and leveraging of existing Si-CMOS process infrastructure needs further efforts and exploration of bottom-up and top-down SiNW fabrication alternatives. Alternatively, some of the most successful bottom-up structures are the CNTs, discovered in 1991 [16]. CNTs have unique mechanical and electrical properties that cannot be matched by any known top-down fabrication techniques. They are highly elastic and with a Young's modulus in the range of 1 TPa [17]. CNT can show ballistic conduction in the range of hundreds of nanometers to micrometers and can carry very high current densities (up to 10^9A/cm^2) [18].

Finally, it is worth noting that the future use of bottom-up SiNW and CNTs in functional NEM devices operating in complex electronic systems will probably require a smart and efficient combination of bottom-up and top-down fabrications.

Hybrid Top-Down/Bottom-Up processing is probably the most efficient and realistic way to build and support collective mechanical computation at nanoscale based on millions of NEM devices. While semiconducting or metal nanowires and nanotubes can be built with controlled dimensions by some selected bottom-up approaches, top-down is used to provide the processing of nanogaps, contacts, and interconnections as well as the interfacing between the nanometer and micrometer worlds (Fig. 3.1).

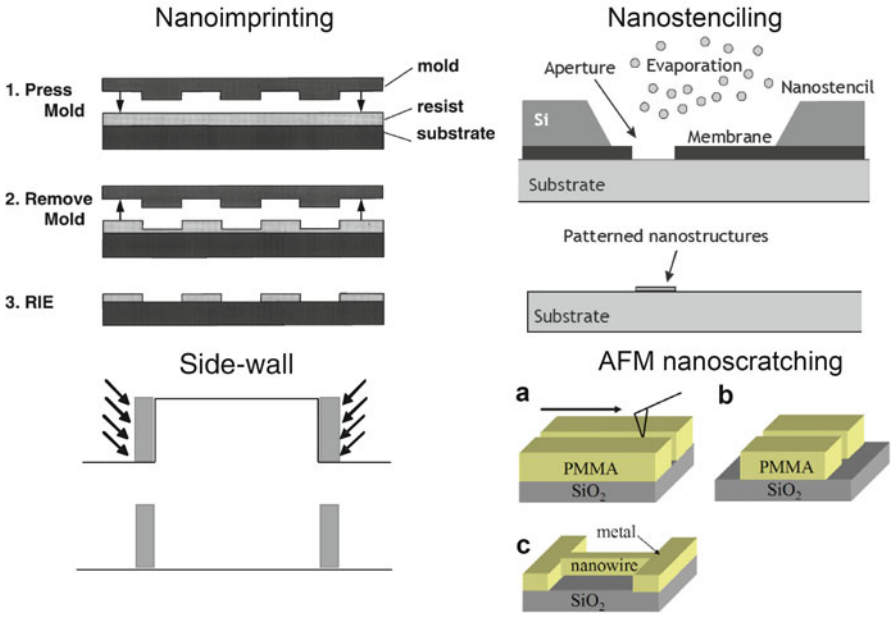

Fig. 3.1 Illustrations of nanoimprinting (*top-left*), nanostenciling (*top-right*), sidewall (*bottom-left*), and AFM nanoscratching (*bottom-right*) as examples of top-down nanofabrication techniques

Overall, although the assembly of bottom-up nanowires and nanotubes into functional NEMS devices shows great progress, this is still dominated by one-by-one device fabrication with a relatively low yield. The top-down fabrication of NEMS is still the most robust and can offer feature size similar with the bottom-up fabrication for very-large-scale integration of NEMS on 200 nm.

In the following, we present three recent examples that illustrate some successful NEMS device processing by top-down and bottom-up techniques.

A first example concerns the hybrid nanoelectromechanical systems (NEMS) array integration strategy combining deterministic bottom-up semiconducting silicon (Si) or metallic rhodium (Rh) nanowire (NW) assembly with conventional top-down microfabrication as illustrated in Fig. 3.2 [19]. Bottom-up assembly is used to position single NWs at lithographically defined locations on a Si chip for fabricating multiplexed arrays. Three separate mechanisms are combined to achieve high-yield NW integration over centimeter-scale chip areas: electric-field forces, capillary forces, and NW liftoff. Arrays of wells are processed in a sacrificial insulating photoresist layer covering metal guiding electrodes defined on the chip surface. Long-range dielectrophoretic forces generated by an alternating voltage in the kilohertz range applied between the guiding electrodes attract the NWs to the surface and align them along the electric field gradient. The NWs that are preferentially aligned in the wells are retained there by capillary forces produced

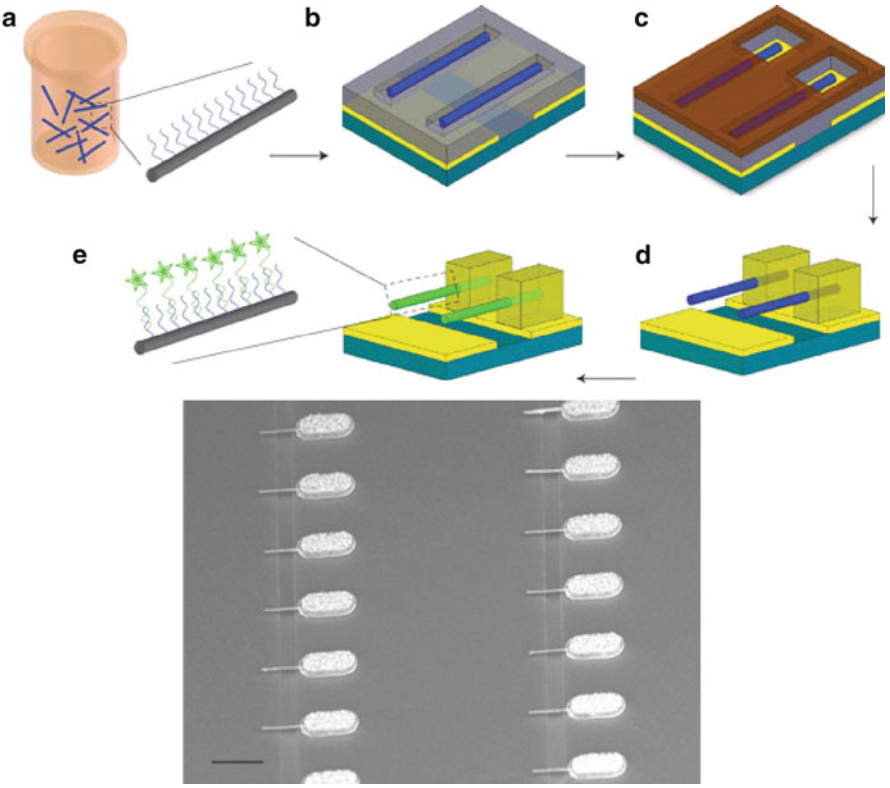

Fig. 3.2 *Left:* Bottom-up fabrication of NW resonator arrays. (**a**) PNA probe molecules are attached to the NWs. (**b**) dielectrophoresis is used to preferentially align single NWs in wells patterned in a sacrificial photoresist layer. (**c**) Individual clamp windows are defined in a second photoresist layer. (**d**) Metal clamps are electrodeposited around the NW tips. The photoresist is dissolved to suspend the clamped NWs and lift off any that are misaligned. (**e**) NWs are exposed to fluorescently label complementary and noncomplementary targets to confirm detection selectivity. *Right:* Rhodium NW resonator arrays with high yield. Reprinted by permission from Macmillan Publishers Ltd: Nature Nanotechnology [19], ©2008

during evaporation of the suspending liquid within each well. The NEMS resonator integration is finalized by electrodepositing a thick metal clamp at one end of the NW aligned in each well and dissolving the photoresist layers.

The second example concerns the realization of a nanoelectromechanical resonator based on a polysilicon suspended cantilever placed on top of a CMOS circuit by using nanostenciling [9]. As CMOS circuits require fabrication compatibility of the postprocessing patterning techniques to avoid any degradation of circuit characteristics during the NEMS definition, nanostenciling appears as a very good candidate to fulfill such requirements. The main steps of the fabrication are depicted in Fig. 3.3. The vibrating nanobeams are defined by nanostencil-deposited Al patterns, and a UV-patterned photoresist mask is used for the releasing aperture.

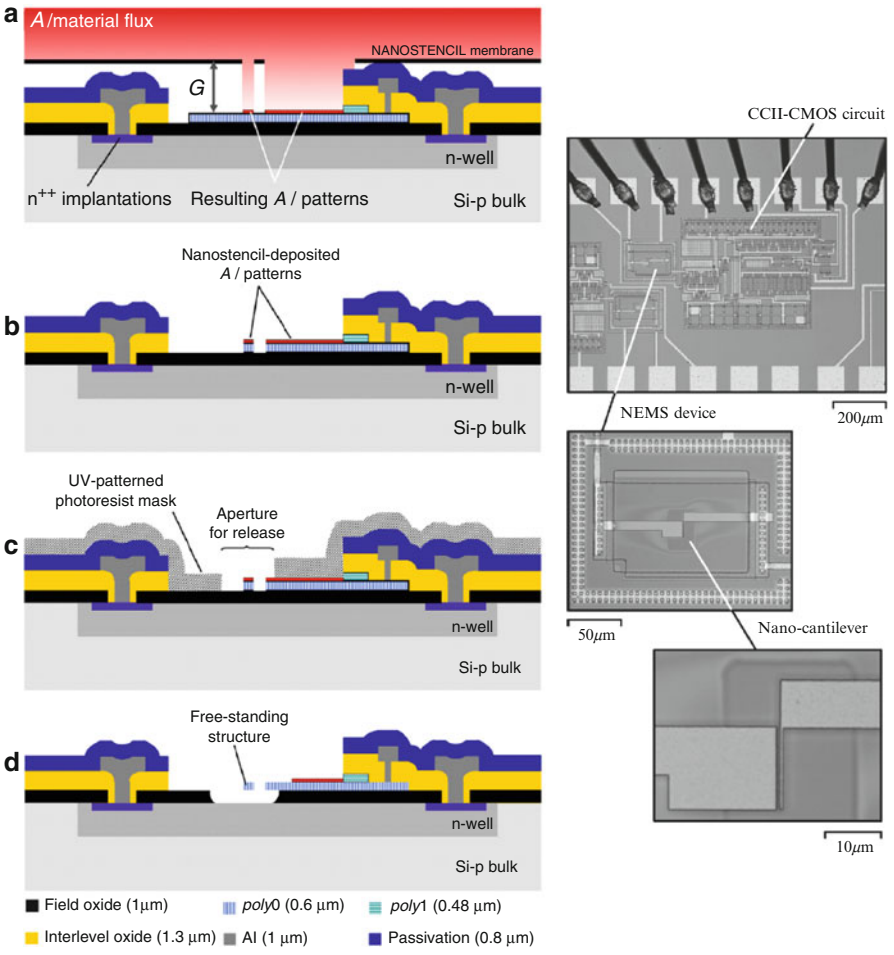

Fig. 3.3 *Left:* Illustration of top-down fabrication of polycrystalline silicon NEMS by nanostencil on top of a silicon CMOS circuitry. ©2008 IOP Publishing. Reprinted with permission from [9]. *Right:* SEM images of resulting hybrid NEMS-CMOS and zoom on the NEM cantilever. ©2008 IEEE. Reprinted, with permission, from [20]

The third fabrication example concerns the fabrication of a CNT NEM structure based on hybrid bottom-up/top-down processing and is depicted in Fig. 3.4 [21]. A 5×5-mm^2 silicon substrate with a 1.5 −μm-thick oxide layer is spin coated with a copolymer of methyl methacrylate and methacrylic acid [P(MMA/MAA)] as a first resist layer followed by PMMA as a second resist layer and baked, Fig. 3.4b. Reference alignment markers are defined by e-beam lithography, Fig. 3.4c. After the development process, 2 −nm Cr (adhesion layer) and 30 −nm Au are deposited followed by liftoff and cleaning. Before SWNTs are adsorbed from a suspension, the surface is functionalized, Fig. 3.4d, with a self-assembled monolayer of 97%

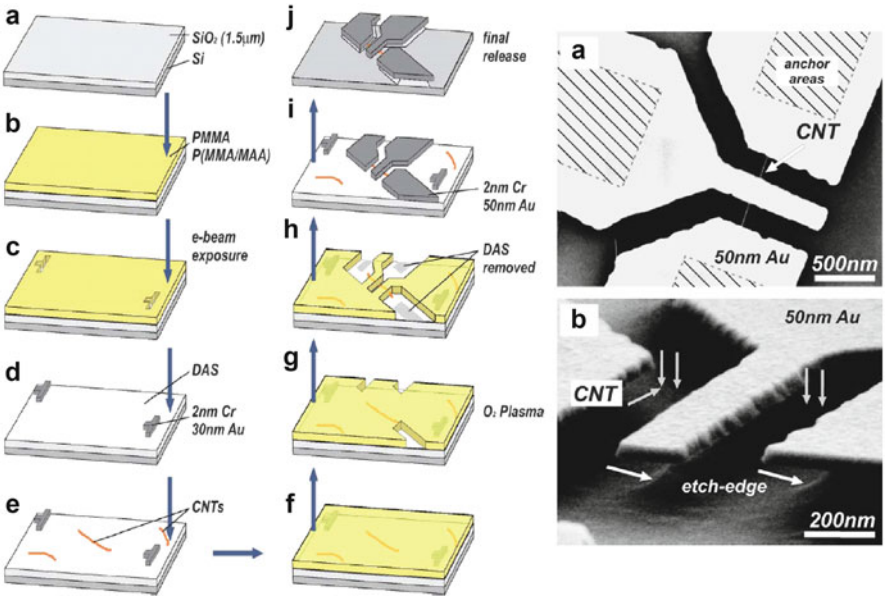

Fig. 3.4 *Top:* Illustration of hybrid fabrication of CNT NEM devices by combining bottom-up and top-down techniques. *Bottom:* SEM images of fabricated NEM device based on suspended CNT. ©2006 IEEE. Reprinted, with permission, from [21]

N-[3-(trimethoxysilyl)propyl]-ethylene-diamine (DAS) to enhance the absorption. The nanotubes adhere to the surface by strong van der Waals forces, which keep them in place during AFM scanning (in tapping mode) and the following e-beam lithography. Deposited reference alignment markers serve to identify the location of each SWNT as is recorded by AFM imaging in tapping mode, Fig. 3.4e. After spin coating again Fig. 3.4f, e-beam lithography is used to structure the individual devices. A short oxygen plasma treatment removes the underlying self-assembled monolayer, Fig. 3.4g as well as all CNTs not covered by PMMA. Further high-precision alignment e-beam and liftoff of Cr as an adhesion layer (1–2 nm) and Au (30–50 nm), Fig. 3.4h,i precede the final HF release, Fig. 3.4j.

3.3 NEM Switches

The main past focus of the semiconductor industry has been on MOS transistor miniaturization to improve performance in terms of speed together with the reduction of the die area and the cost. This has been a successful approach during the *happy scaling era*, but as the transistors have been scaled to dimensions much below 100 nm, many other challenges have appeared. Most importantly, the off-state leakage current, I_{off}, has increased exponentially as the transistor threshold

voltage, V_{th}, was reduced in order to achieve increased transistor speed. Today that static (leakage) power dissipation comprises a large fraction of the total power consumption of a chip, limiting voltage scaling in nano-CMOS.

The power issue is fundamentally due to the nonideal switching behavior of a transistor in the subthreshold region. An ideal switch has an infinitely abrupt transition between the off and the on states, as the gate control voltage is varied. It has zero subthreshold swing defined as $SS = [d(logI_D)/dV_{GS}]^{-1}$. Such an ideal switch would allow for a very low operating voltage (i.e., a low supply voltage V_{dd} of $\sim$100 mV) and hence very low static and dynamic power consumptions. Unfortunately, the subthreshold swing of a MOSFET is limited by Boltzmann statistics to be no less than 60 mV/dec at room temperature and so needed voltage must be much larger than 60 mV in order to allow for reasonably large I_{on}/I_{off} ratios (for instance, $I_{on}/I_{off} > 10^5$ needs a voltage swing $SS > 300$ mV in MOSFET operating at room temperature). Note that for many conventional sub-100 −nm MOSFETs, the value of SS is usually larger than the 60 mV/decade fundamental limit. Multiple-gate MOSFET structures offer improved electrostatic control of the channel potential and therefore can achieve SS values close to the limit of 60 mV/decade. Therefore, more ideal switching devices based on *other physical mechanisms* are needed to overcome the power issue. Various solutions have been proposed in order to break the 60 mV/decade subthreshold swing limit: impact ionization devices, tunnel FETs, negative capacitance FETs, and electromechanical devices. Each of them has their advantages and disadvantages. The nanoelectromechanical (NEM) switch or relay, and the hybrid NEM field-effect transistor (NEM-FET) presented and discussed in this chapter are two key examples of electromechanical device candidates for abrupt switching [22].

3.4 NEM Relays

A NEM relay can be designed similarly to a MOSFET as a three- [23] or four- [24] terminal in-series switch where an air-gap separates one movable conductive electrode and one fixed conductive electrode, see Fig. 3.5a,b. Recently a two-terminal microelectromechanical diode switch was also reported [25], Fig. 3.5c. In Fig. 3.5a, these two electrodes can be considered analogous to the "source" and "drain" electrodes of a MOSFET. In the off state, the air gap between source and drain electrodes provides zero I_{off} (or zero conductance). As the "source" and "drain" electrodes are brought into physical contact via an actuation mechanism (very often, electrostatic actuation) using the control offered by the "gate" electrode, the conductance increases abruptly with near-zero subthreshold swing, see Fig. 3.6. Device operation is based on electrostatic actuation, and the current-voltage characteristics usually feature a hysteresis characterized by a pull-in voltage, V_{pi}, which is larger than the pull-out voltage, V_{po} (Fig. 3.6). In addition to the relay operation, the NEM switch hysteresis can be exploited for one-transistor memory functions. The actuation direction can be out-of-plane or in-plane. Three-terminal NEM relays

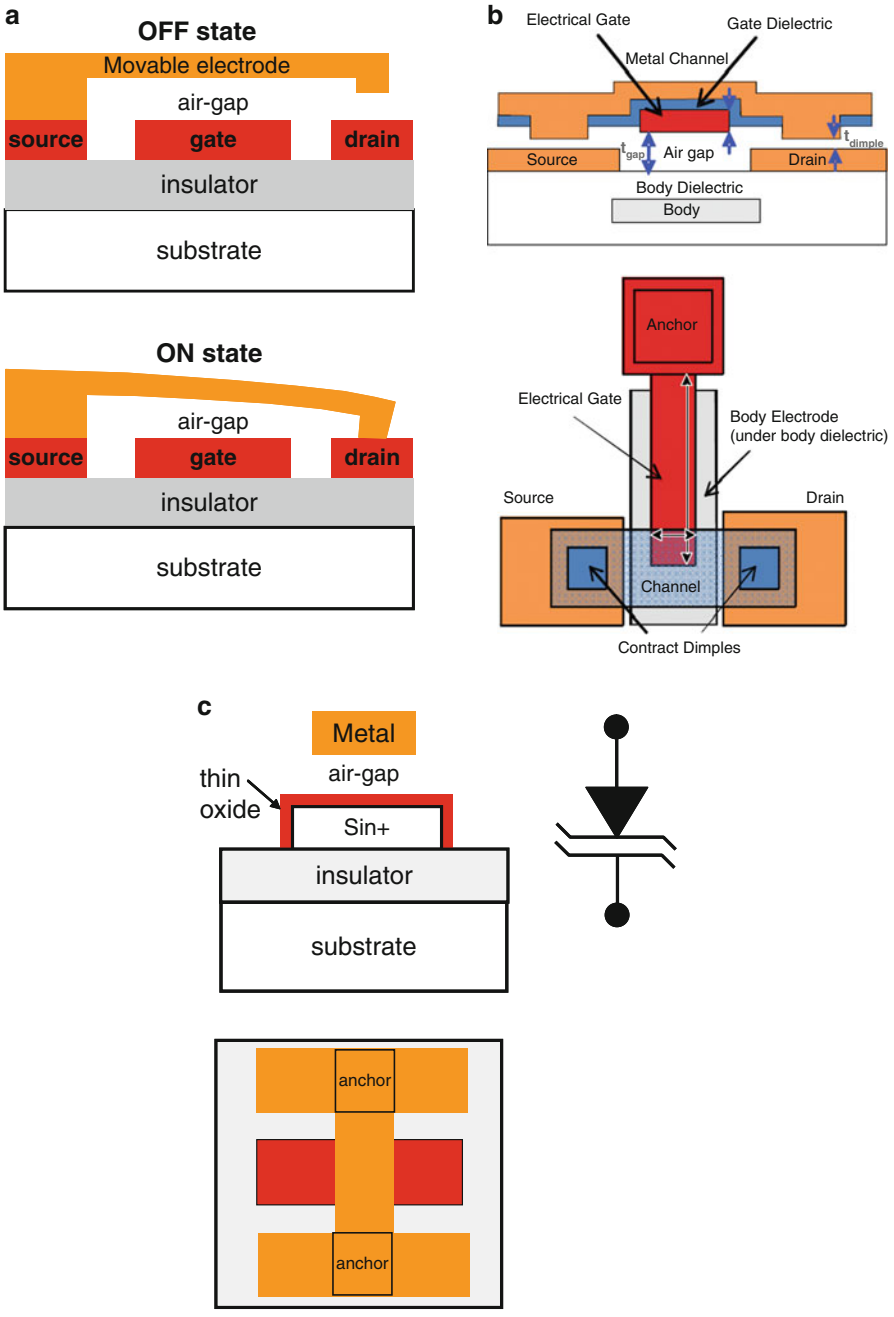

Fig. 3.5 (**a**) Three-terminal NEM switch analogous to MOSFET device; OFF and ON states defined by electrostatic actuation using gate potential, (**b**) Four-terminal NEM switch (©2008 IEEE. Reprinted, with permission, from [24]) and (**c**) Two-terminal NEM tunneling diode switch

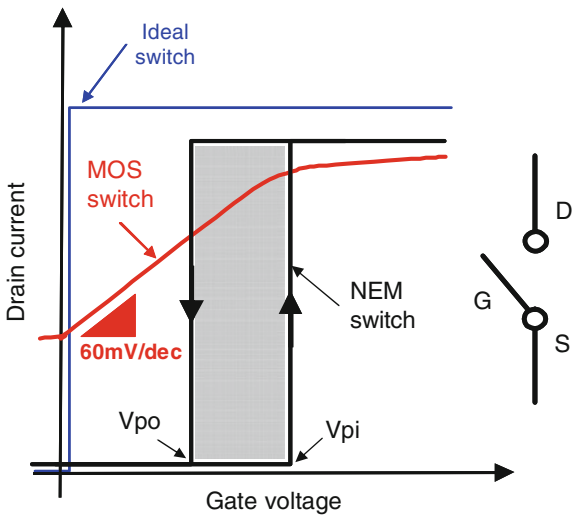

Fig. 3.6 Qualitative representation of $I_D - V_G$ characteristics of three-terminal switches: ideal, MOS, and NEM. The main advantages of in-series NEM switch are the ultralow I_{off} (practically, zero static current/power) and the quasi-infinite abrupt transition between off and on states

can be interconnected as the MOSFETs and can be used to design logic families [23, 24]. Four-terminal relays can be used to design more complex functionality so that a given logic function can be implemented with a smaller number of devices compared with CMOS implementation, compensating in terms of number of devices the larger device area needed for a NEM relay compared to a solid-state transistor.

It is well known that any micro- or nanoelectromechanical relay has an associated problem of contact wear due to opening and closing of the mechanical contacts. The mechanical opening/closing and the high current density through the resulting imperfect contact can naturally limit the lifetime of such mechanical relay. On the other hand, although solid-state semiconductor switches do not have such contact problem (resulting in longer operational lifetimes), they have other intrinsic limitations such as high leakage (I_{off}) current and a significant dependence and increase with the temperature of their leakage currents. Thus, the major advantages that NEM switches can offer compared to their solid-state counter parts are (a) ultralow standby power (due to air-gap isolation in off state and abrupt switching between off and on states, see Fig. 3.4), (b) reduced dependence of I_{off} on environmental conditions such as temperature, and (c) excellent high-frequency characteristics (such as isolation, insertion loss).

Overall, NEM switches have a clear potential for low dynamic energy dissipation, which can be reduced down to the aJ. Figure 3.7 shows the energy-throughput tradeoffs in CMOS and NEM relay adders [24]. Here the adders have been designed to drive a load capacitance of either 25 fF or 100 fF, and in order to isolate the energy dissipation of the adder itself and enable comparisons with respect to driving different loads, the plot shows the energy per operation (E_{op}). A reduction of about two orders of magnitude in E_{op} is obtained with the price of a slower operation. Applications requiring throughputs of the order of 50–100 MOPS or less would appear to benefit from such NEM relay technology [24].

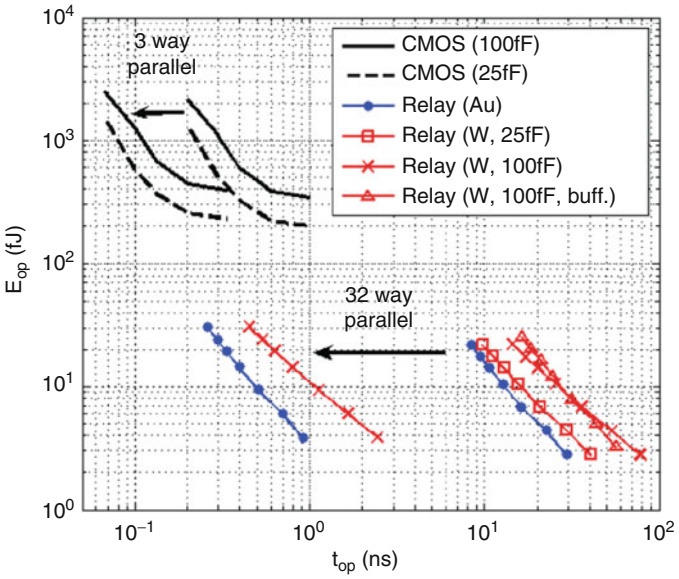

Fig. 3.7 Comparison of energy-throughput trade-offs in CMOS and NEM relay adders. ©2008 IEEE. Reprinted, with permission, from [24]

Furthermore, in case of CNT relays, one can also foresee NEM relays with significantly increased current density and power handling, drastically extending the reliability and lifetime of mechanical contacts and avoiding or limiting electromigration phenomena. Given the extremely small cross section of CNTs (thus a limited current per individual tube), the integration of such concepts in electronic circuits requires the design and fabrication of arrays of CNT switches. While Fig. 3.6 qualitatively compares the absolute values of the current, in-series CNT relays might achieve higher current density compared to ultrascaled MOSFETs. However, in order to drive currents as high as hundreds of μA or mA, many parallel tubes should be connected, which gives a motivation for the concept of arrays of parallel-connected CNT NEM switches. Various reports show that the main weakness of NEM switches is their switching speed: a realistic lower limit for the delay related to beam movement from the off position to the on position is of the order of 1 ns [23].

NEM relays can exploit CMOS-like processing or can be cointegrated with silicon CMOS. Foreseen applications of a NEMS relays hybridized with CMOS consist of power gating high-performance CMOS circuits [26] and configuration of CMOS FPGAs [27]. Another key application opportunity of NEM relays resides in their combined improved energy efficiency and potential for 3D integration (due to low-temperature processability), both of which may allow increased functional density for a given substrate real estate. This is especially interesting for memory applications where NEM-intrinsic hysteresis (Fig. 3.6) and sticking can be turned into practical advantages.

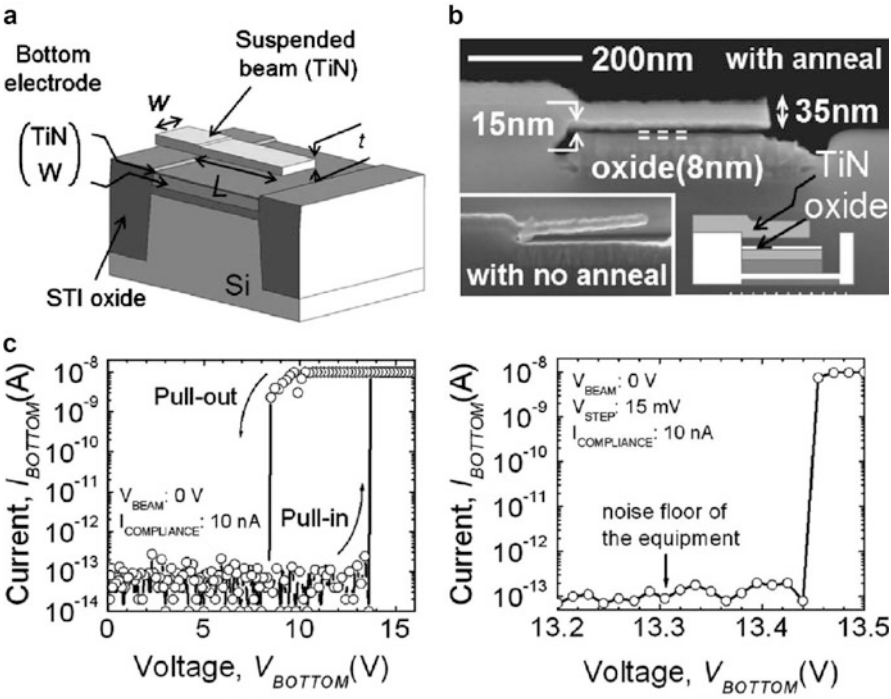

Fig. 3.8 (**a**) NEM cantilever switch, (**b**) cross-sectional view of 35 −nm-thick TiN cantilever suspended over a 15 −nm air gap (**c**) I-V experimental plots of the fabricated NEM cantilever switches showing abrupt switching and quasi-zero off current (noise floor of the measurement equipment also shown). Reprinted with permission from [28]. ©2008 American Institute of Physics

NEM relays can be fabricated by lithographic top-down or bottom-up approaches using CNT or nanowire beams, as described in the previous section. There are many successful demonstrations using top-down approaches at the micron scale with reliable operation up to 10^9 cycles [27]. The nanoscale contact reliability is particularly challenging since logic circuits would require the relays to operate without failure over $\sim 10^{16}$ cycles in "hot switching" conditions.

Another critical process step is the beam release (air-gap formation with nanometer feature size), which is realized by the etching of a sacrificial material, such as oxide, polyimide, or silicon. The smallest actuation gap demonstrated so far for a functional NEM memory structure fabricated using a top-down approach is 15 nm, as shown in Fig. 3.8 [28]. The device consists of a vertically actuated 2-terminal NEM switch featuring a 35 −nm-thick TiN cantilever beam that is 300 nm long and 200 −nm-wide. The pull-in voltage was in the order of 13.45 V and the off-state current is quasi-zero with a subthreshold swing practically zero. The fabricated cantilever-type NEM switch showed an abrupt switching with less than 3 mV/decade, and an on/off current ratio exceeding 10^5 in ambient air.

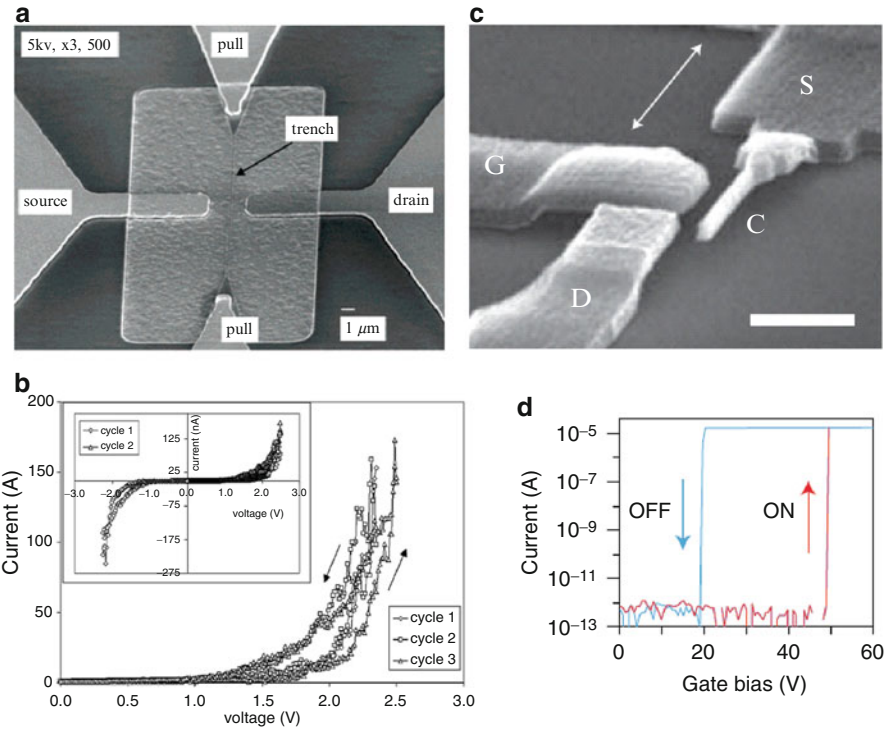

Fig. 3.9 Bottom-up NEM CNT switches: (**a**) individual CNT suspended over gap and (**b**) its current-voltage characteristics showing less-than-2V operation. Reprinted with permission from [29]. ©2006 American Chemical Society. (**c**) All-SWCNT NEM relay and (**d**) its static current versus voltage characteristics, showing abrupt off–on transitions, quasi-zero I_{off}, and hysteretic characteristics. The pull-in voltage is much higher for the all-SWCNT switch because of its much larger air gap. Reprinted by permission from Macmillan Publishers Ltd: Nature Nanotechnology [30], ©2008

The endurance property could be improved by using insulating layer with high melting temperature and strong hardness, such as Al_2O_3 and HfO_2.

The remarkable electromechanical and electrical properties of carbon nanotubes (CNTs) make them excellent candidates for the design of NEMS. In [29] NEM structures based on a single-walled nanotubes (SWNTs) with a diameter of the order of 1–3 nm suspended over shallow trenches (air gap of the order of 20 nm) in a SiO_2 layer, with a Nb pull electrode beneath, were reported. The nanotube growth is performed on-chip using a patterned Fe catalyst and a methane chemical vapor deposition (CVD) process at 850 °C. The corresponding SEM micrograph of a finished device is shown in Fig. 3.9a. The underlying Nb pull electrode in the nanotrench contacted by the Au/Ti electrodes is labeled "pull" in the same figure. Figure 3.9b reports the switching characteristics. At 3.5 V, the current begins to rise rapidly to ∼250 nA at ∼4.5 V. The I_{on}/I_{off} is larger than 10^4, implying well-defined

off and on states. The authors explain the measured currents based on tunneling mechanisms and hysteretic behavior are well observed in the same work.

Overall, the works on NEM switches agree on the fact that, when the contribution from van der Waals forces is ignored, the switch pull-in voltage, V_{PI}, to first order, is calculated using:

$$V_{PI} = \sqrt{\frac{8k_{eff}g^3}{27\varepsilon_0 wL}}, \tag{3.1}$$

where g is the air gap, ε_0 is the effective permittivity, w is the beam width, and L is the length. The effective spring constant, k_{eff}, depends on the Young modulus, the inertia modulus, and the clamping design of the suspended structure.

One of the limitations of NEM switches based on individual structures is the limited current level that usually stays below 1–10 μA for a single NW or CNT. Therefore, there is a practical necessity of assembling parallel arrays of NW or CNT switches. Recently, the use of "CNT wafers" rather than individual CNTs has been demonstrated as a very promising approach for fabricating high-density, highly controllable CNT-based NEM relays and logic circuits [30]. Hayamizu et al. [30] have proposed a hierarchical assembly of carbon nanotubes into closely packed and highly aligned three-dimensional wafer films from which a wide range of complex and three-dimensional nanotube structures were lithographically fabricated. They have fabricated an all-SWCNT CNT relay (Fig. 3.9c) composed of a 170 −nm-wide cantilever connected to a 3D electrode associated with two 3D nanotube source and gate electrodes arranged in predefined locations. All device parts are made from the same SWNT film and located with high precision to achieve operational devices. The electromechanical switching is demonstrated in Fig. 3.9d; the high pull-in voltage (∼50 V) is due to the large air gap. This CNT switch can pass currents well beyond 100 μA (e.g., 270 μA is experimentally demonstrated). Future advancement on this thin-film CNT approach could create controlled CNT-NEMS devices and offer to the circuit designer novel applications opportunities.

3.5 Nanoelectromechanical Field-Effect Transistors as Abrupt Hysteretic Switches

The first electromechanical transistor, called resonant gate transistor, was proposed by Nathanson in 1967 [31]. The main functionality of the device was not related to its mechanical electroswitching behavior but rather to the exploitation of the mechanical resonance of a movable gate and the possible use in electronic oscillators. However, because at the respective time the MEMS technology was only in a very early stage, further developments were very limited and later reports on transistors with suspended gate were dedicated to applications in gas sensing where a suspended gate design permits the access of gas species to the inner surface of the gate which acts as chemical-sensitive layer [32]. Recently, so-called suspended-gate

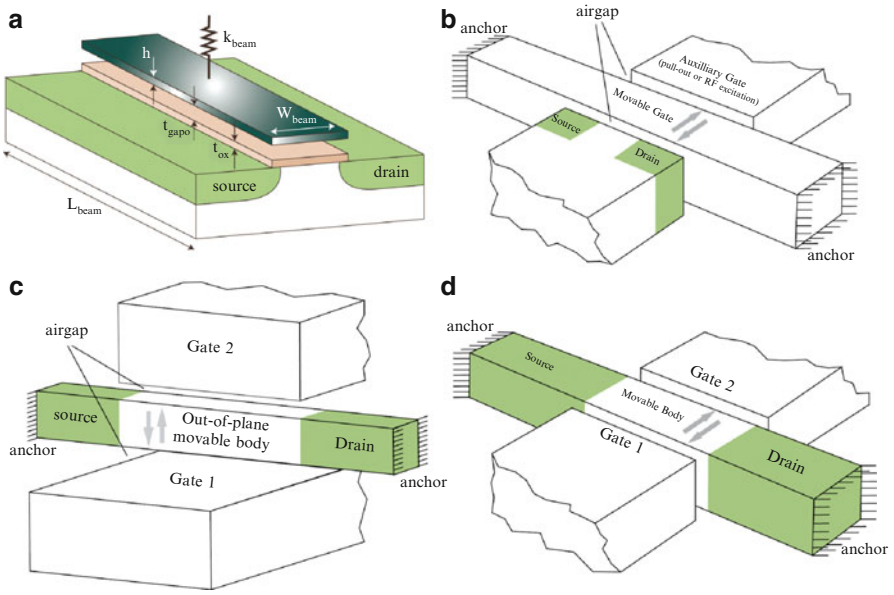

Fig. 3.10 Various technological implementations of NEM-FETs: (**a**) out-of-plane movable gate. (**b**) in-plane movable gate and auxiliary gate electrode. (**c**) out-of-plane movable body and double-gate configuration. (**d**) in-plane movable body and double-gate configuration

FET (SG-FET) devices [33–36] have been proposed with main emphasis on their ability to offer solutions for ultralow-power logic, power management, and capacitorless memory devices by taking advantage of the properties of the hybrid MEM-solid-state transistor.

A Micro-/nanoelectromechanical field-effect transistor (MEM- or NEM-FET) combines the features of a mechanical relay and a MOS transistor. It has a conductive movable part (movable gate or movable body) and a solid-state semiconductor part (transistor channel between source and drain regions), see Fig. 3.10a–d, that operate to couple the mechanical movement with the formation of the inversion or accumulation channels at the gate-insulator/semi-conductor interface. In a MEM-FET, the effective gate dielectric thickness is dynamically modulated by the applied gate voltage. The resulting applied electrostatic force moves the suspended gate or the suspended body and changes the air gap and gate-to-channel capacitance. MEM-FETs with movable gate-electrode dimensions on the order of 10 µm, or movable body dimensions on the order of 1 µm, and air gaps in the range from 100 to 200 nm and actuation voltages in the range of a few volts have been successfully demonstrated [34].

The MEM-FET (called SG-FET in the original publications [34, 35]) device architecture with out-of-plane (vertical) suspended gate is depicted in Fig. 3.11a and its experimental behavior of Fig. 3.11b. The movable part determines the state of the gate capacitance C_{gg} (either low or high), which determines the threshold voltage.

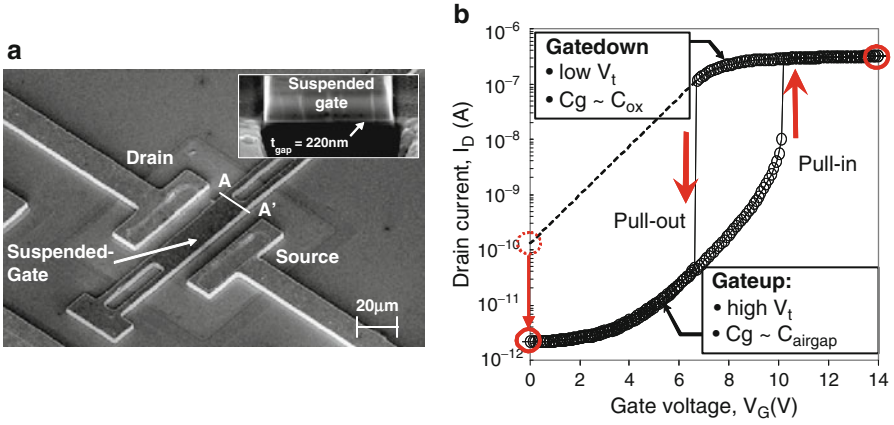

Fig. 3.11 (**a**) SEM image of fabricated SG-FET with metal gate and airgap of 220 nm. ©2006 IEEE. Reprinted, with permission, from [34]. (**b**) Drain current, I_D, versus gate voltage, V_G, hysteretic characteristics showing pull-in and pull-out voltages of the SG-FET switch

Therefore, the SG-FET is a dynamic threshold voltage (V_t) device, with a high V_t in the off state (providing a very low I_{off}) and a low V_t in the on state (providing a high I_{on}). In the SGFET, when the gate is pulled in, the gate capacitance increases abruptly, and so do the surface potential, the charge density, and the drain current. As can be seen from Fig. 3.11b, an abrupt transition (with an experimental swing as low as 2 mV/decade, reported in [34]) occurs as the gate voltage is increased beyond the point when an imbalance between the electrostatic and mechanical spring-restoring forces is reached so that "pull-in" of the mobile gate occurs. A similarly abrupt transition from the low-V_t state to the high-V_t state occurs as the gate voltage is decreased back to 0 V, when the electrostatic force is reduced to be equal to the spring restoring force so that "pull-out" occurs.

Note that in SG-FET the current path is always in the semiconductor part, so that this device is not subject to contact reliability issues specific to mechanical relays. However, NEM-FETs could be subject of oxide charging, resulting in shifts of the pull-in voltages; this effect can be minimized by appropriate design or even exploited for nonvolatile memory applications. Memory applications of suspended gate MOSFET have been reported based on experiments [35] or simulations [36]. Recently, the experimental characteristics of a fabricated in-plane double-gate movable-body SOI MOSFET (Fig. 3.12a) with lateral nanogaps have been reported showing similar abrupt transition and hysteretic behavior controlled by the lateral gate voltages [37]. The reported hysteresis (Fig. 3.12b) can be exploited to define stable high and low states for memory applications, using the combined biases of the two lateral gates.

For logic circuit applications of NEM-FETs, the pull-in should occur at a gate voltage less than the (high) off-state V_t to achieve substantial subthreshold leakage power savings for a given on-state current specification [38]. Furthermore,

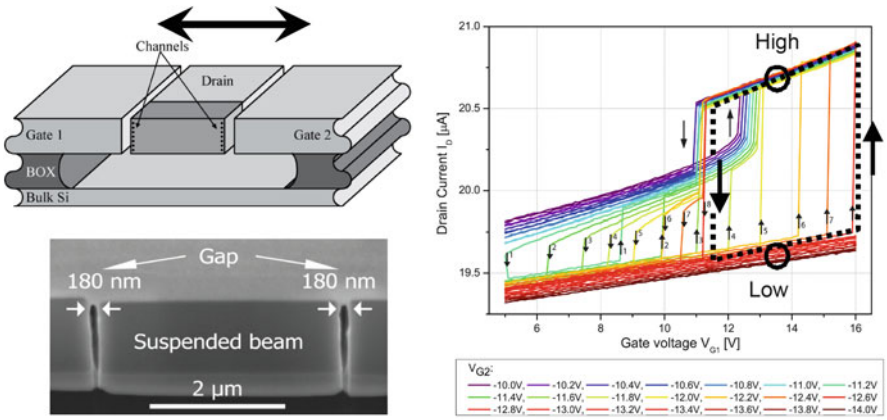

Fig. 3.12 (**a**) Drawing and SEM cross-sectional view of movable body FET on thin SOI, (**b**) experimental characteristics of a fabricated double-gate in-plane movable body FET. ©2008 IEEE. Reprinted, with permission, from [36]

pull-out should occur at a gate voltage greater than 0 V for proper circuit operation. Comprehensive analytical models to predict NEM-FET device characteristics, to predict the pull-in and pull-out with equations specific to the hybrid NEM-FET devices, and, finally, to enable accurate circuit simulation and design have been proposed in [38, 39]. Supposing that the switching occurs in weak inversion, the pull-in and pull-out voltages can be calculated as:

$$V_{pi} = V_{FB} + \overline{\gamma} \cdot \sqrt{\psi_{pi}} + \psi_{pi}, \tag{3.2}$$

$$V_{po} = V_{FB} + \overline{\gamma} \cdot \sqrt{\psi_{po}} + \psi_{po}, \tag{3.3}$$

where

$$\overline{\gamma} = \gamma \cdot \left(1 + \frac{C_{ox}}{\varepsilon_{gap}/x_{pi}}\right), \tag{3.4}$$

and

$$\gamma = \frac{\sqrt{2\varepsilon_{Si} q N_A}}{C_{ox}}, \tag{3.5}$$

with ψ_{pi} and ψ_{po} being the surface potential values at pull-in and pull-out, respectively. These equations reduce to the well-known equations on MEM switches when the semiconductor substrates transform in an ideal conductor ($N_A \to \infty$) leading to $\psi_{pi} \to 0V$ and $x_{pi} \to 2t_{gap0}/3$.

The measurement of nanodisplacements based on in-plane suspended gate MOSFET detection fully compatible with front-end CMOS processes has been demonstrated in [40].

Although truly nanometer-scale NEM-FETs have not been experimentally reported to date, successful demonstrations of micron-scale devices together with simulations of scaled NEM-FETs with calibrated models indicate that NEM-FETs

hold an interesting potential for low-standby-power applications. Moreover, the use of NEM-FET as resonating devices in analog/RF (detailed in next section) and sensing applications appear to be the most promising for the future.

3.6 Nanoelectromechanical Resonators

3.6.1 Principles, Opportunities and Limits

Historically, some of the first human-made resonators have been musical instruments and chronographs operating from Hz to kHz frequencies. The increasing recent interest in wireless applications [41–43] has generated new efforts to miniaturize and integrate the off-chip passives needed for filtering and frequency generation in communication circuits. Among various off-chip components of a communication transceiver, the quartz crystal used in the reference oscillator appears to be the most difficult to miniaturize and integrate on chip while preserving its off-chip performance. Its high quality factors, Q ($>10^4$–10^5), simultaneously combined with unique temperature stability (uncompensated, much better than 1 ppm/°C) have been unbeatable by any IC technology until very recently. The quality factors of integrated LC tanks are severely limited by the poor-quality-factor values of integrated inductors and capacitors, ranging in best cases from tens to hundreds. Even with more sophisticated techniques like suspended postprocessed and suspended L and C using advanced Cu interconnect layers [44], the values of Q stay limited at RF frequencies. This is naturally related to losses coming from their series resistances, substrate coupling, and supplementary capacitance parasitics related to the particular device design. Moreover, the technologies proposed for optimizing the performances of inductors and capacitors are different and could frequently require divergent efforts and complex processing.

As a consequence, the most promising solutions to miniaturize reference oscillators with uncompromised quality factors [41] are related to the classes of vibrating devices, based on IC-compatible micromachining processes and materials like semiconductors, polysilicon, or metals. Among the most promising of these vibrating structures are the capacitively transduced micro- and nanoelectromechanical (M/NEM) resonators and the thin-film bulk acoustic piezoelectric resonators (FBARs). The principle of M/NEM resonator with capacitive transduction is depicted in Fig. 3.13a. When an ac signal is applied on the fixed electrode and capacitively coupled to the beam at an applied signal frequency matching the mechanical resonance frequency, f_0, of the clamped–clamped beam displacement, thus the output capacitive current is maximum and a maximum of the transfer function, bandpass-like (Fig. 3.13b), is obtained. The resonance frequency is expressed as:

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k_{eff}}{m_{eff}}}, \tag{3.6}$$

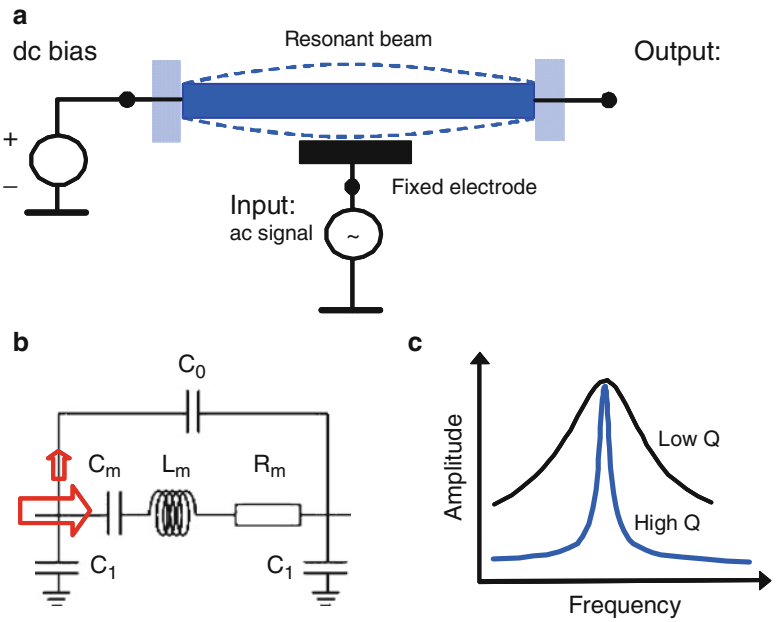

Fig. 3.13 (**a**) Principle of clamped–clamped beam resonator, (**b**) equivalent circuit of the micro-electromechanical (MEM) resonator: series *R-L-C* circuit and parasitic capacitances, C_0 and C_1

where k_{eff} is the effective stiffness and m_{eff} the effective mass. The equivalent circuit of the resonator is shown in Fig. 3.13c. It essentially consists of a series R_m, L_m, C_m circuit, developed by using an electrical-to-mechanical analogy, as follows: (a) the motional inductance, L_m, corresponds to the resonator mass, m; (b) the motional resistance, R_m, to the damping (coefficient), b; and (c) the inverse of the capacitance, $1/C_m$, to the stiffness, k. The equivalent circuit components can be expressed as functions of the mechanical parameters of the structure, the applied DC voltage, and the capacitance derivative with respect to displacement as follows:

$$L_m = \frac{m_{eff}}{2V_{DC}^2(\partial C/\partial x)^2}, \tag{3.7a}$$

$$C_m = \frac{2V_{DC}^2(\partial C/\partial x)^2}{k_{eff}}, \tag{3.7b}$$

$$R_m = \frac{\omega_0 m_{eff}}{2QV_{DC}^2(\partial C/\partial x)} = \frac{k_{eff}g_0^4}{\varepsilon_0^2 A^2 \omega_0 Q V_{DC}^2}. \tag{3.7c}$$

Of major importance is the value of the motional resistance that should be as small as possible. Additionally, Fig. 3.13 suggests the importance of the values of

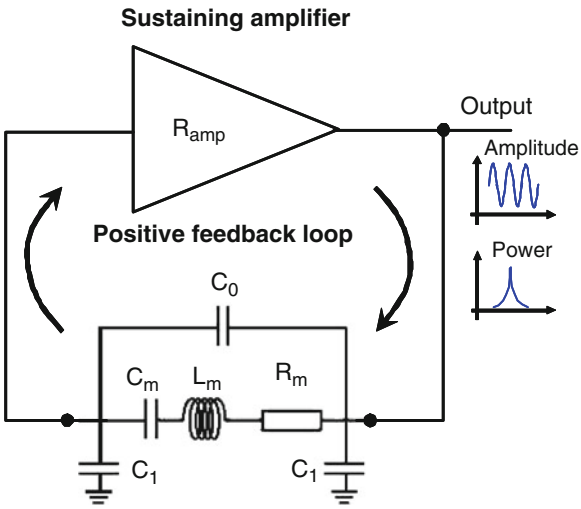

Fig. 3.14 Typical MEM oscillator configuration using a sustaining amplifier and a MEM resonator (depicted by its equivalent circuit) in positive feedback loop configuration

capacitances C_0 and C_1 for the proper exploitation of a resonator; at resonance the value of R_m (equivalent impedance of the series circuit) should be much smaller than $1/(\omega C_0)$. In practice, with k_{eff} and Q quasi-fixed and C_0 not easily scalable, this requires a severe reduction of R_m by using nanogaps (low value of g_0) or by applying a higher DC bias (V_{DC}). The values of C_0 and C_1 are also crucial for the design of the oscillator circuit. Very high quality factors of M/NEM resonators reflect the possibility to have very low energy losses compared to the energy stored in these structures. Figure 3.14 shows the typical design of an M/NEM oscillator combining a sustaining amplifier and a resonator to provide a positive feedback loop.

The general trend being to increase their resonance frequency from MHz to frequencies exceeding the GHz, the needs in terms of very small, very stiff, and low mass systems have pushed MEM into NEM designs of such resonators, exploiting materials with high Young modulus like silicon, carbon nanotubes, and diamond. However, despite the obvious potential given by NEM resonators in terms of higher resonance frequencies, their ability to conserve a high-Q when the losses cannot be scaled anymore can be questioned as it depends on the amount of energy stored in the structure (thus, its geometrical size). For instance, Fig. 3.15 reports the general observed quasi-linear trend of having smaller quality factors in fabricated resonators when their volume is aggressively scaled down. The limits of quality factor in NEM resonators based on one-dimensional structures such as nanowires and nanotubes have been discussed in [43]. The main energy dissipation mechanisms in NEM resonators resulting in low quality factors Q (100's to 1,000's) are gas friction, clamping loss, and surface loss. However, recent experiments based on microwave spectroscopy on vibrating molecules showing ultra-narrow spectral width with high-Q factors suggest that this scaling trend of smaller Q at smaller size is probably not a fundamental one.

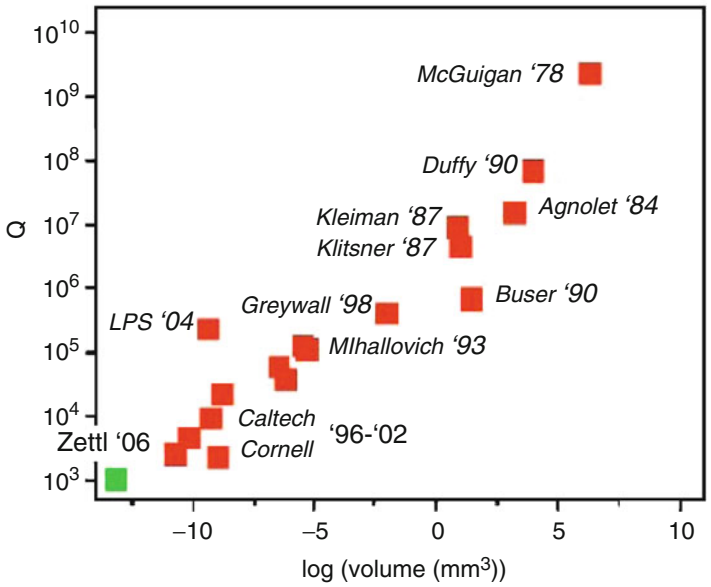

Fig. 3.15 Quality factor versus volume of the MEM resonator (both in log scale); various experimental data. ©2006 IEEE. Reprinted, with permission, from [43]

Another important question concerning the limits of NEM resonators is how the (in)stabilities of these resonators scale with dimensions. For instance, when a resonator device becomes very small, the effects of fluctuations in numbers of photons, phonons, electrons, and adsorbed molecules can significantly affect the noise characteristics [45]. It was suggested that at nanometer scale the frequency noise due to temperature fluctuations is likely to limit their performance and usability. Figure 3.16a reports the effect of the volume scaling of a clamped–clamped silicon beam resonator on the root-mean-square (RMS) temperature fluctuations and the corresponding RMS fractional frequency fluctuation, assuming a frequency versus temperature coefficient of −10 ppm/K. In Fig. 3.16b, the comparison of the phase noises due to the temperature fluctuations caused by the combined effects of radiation and conduction in silicon NEM resonators versus a quartz reference is shown. The temperature-fluctuation-caused noise of the Si resonators are significantly higher than that of the quartz resonator, for example, by ∼70 dBc/Hz at 1 Hz when comparing 10 MHz Si and quartz resonators. However, when connecting N resonators in series, a $10logN$ reduction in both flicker and white phase noise floor levels was observed [46]; therefore, one possible solution to reduce noise in ultrascaled resonant devices could be to exploit arrays with a large number of components.

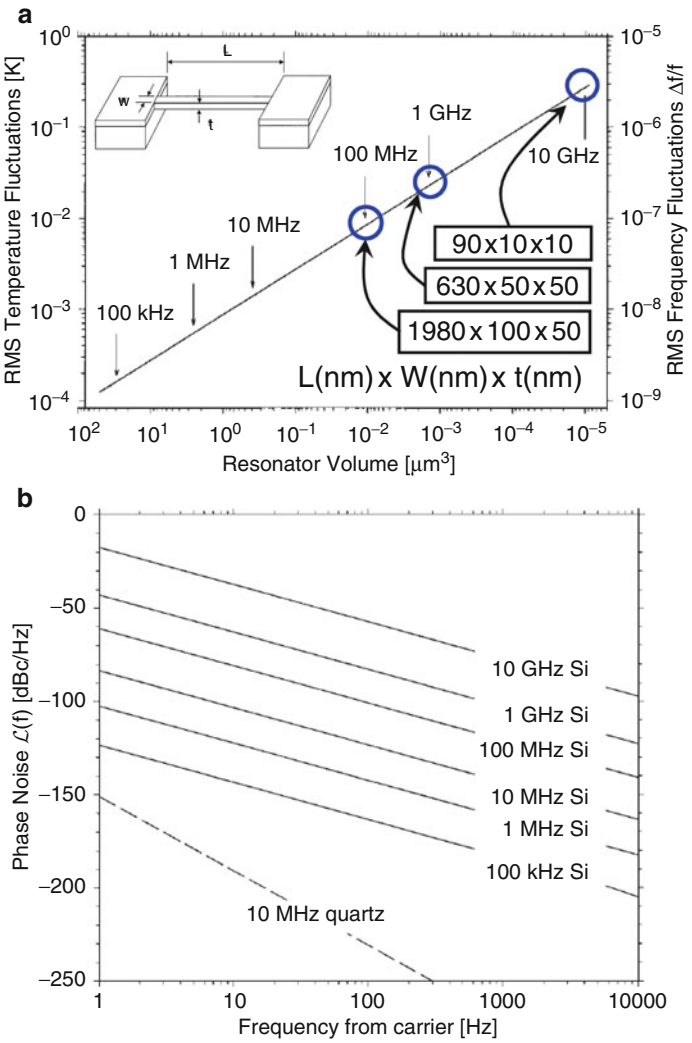

Fig. 3.16 RMS temperature fluctuations and corresponding fractional frequency fluctuation as a function of resonator volume, assuming a frequency versus temperature coefficient of −10 ppm/K, for different resonator beam dimensions. ©1999 IEEE. Adapted, with permission, from [45]

3.6.2 High-Frequency Silicon Nanowire and Carbon Nanotube Resonators

The availability of high-frequency resonators, with resonance frequency possibly higher than 100 MHz to 1 GHz are highly desired for on-chip implementations of the preselect or channel-select filters needed in the RF front-ends of wireless

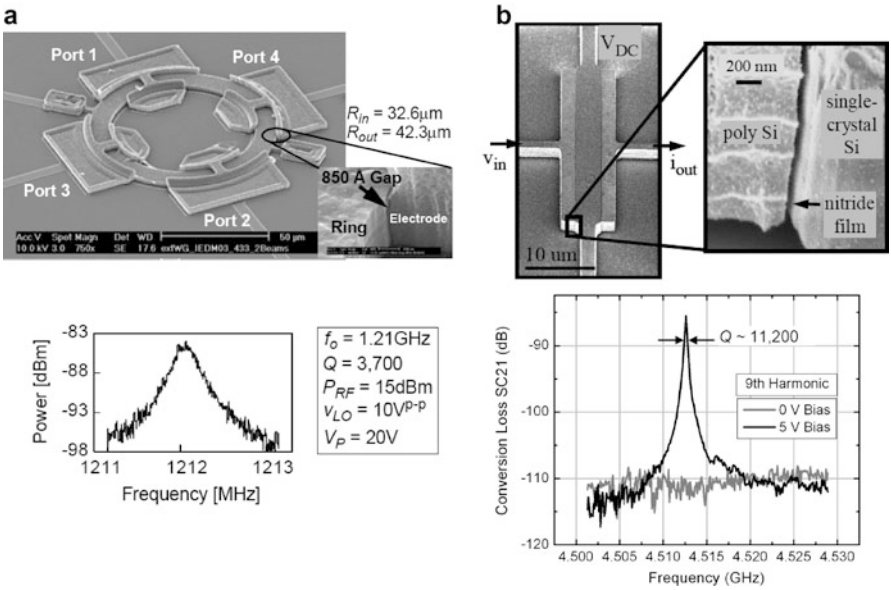

Fig. 3.17 (**a**) 1.2 −GHz polysilicon ring resonator with 78 −nm air gap, ©2003 IEEE. Reprinted, with permission, from [47]. (**b**) 4.41 −GHz resonator with dielectric actuation and piezoresistive sensing, ©2007 IEEE. Reprinted, with permission, from [49]

communication applications. The availability of miniaturized in-IC RF filters based on M/NEM resonators is expected to simplify the realization of future multiband reconfigurable wireless communicators, requiring many of such RF filters compared to today's communication architectures [47].

At micrometer scale, recent successful demonstrations of very-high-frequency resonators are the extensional wineglass resonators with frequencies ranging from 400 MHz to 1.5 GHz (and $Q > 3{,}700$) [47] and the dielectrically actuated and piezoresistively sensed 4.41 −GHz silicon bar resonator exploiting internal dielectric actuation [48]. Figure 3.17a shows the polysilicon ring resonator structure with electrode-to-resonator spacing of 78 nm proposed in [47], capable of achieving high resonance frequency (>1.2 GHz), high quality factor (>3,700) and low impedance (∼200 kΩ at 10 V applied voltage). In [48, 49], the piezoresistive sensing of a capacitively actuated resonator exceeding 4 GHz with $Q > 8{,}000$ and using the 9th harmonic longitudinal mode, is implemented (Fig. 3.17b). In this design and operation, the resonator geometry and frequency scaling are more favorable than for capacitive sensing, and additionally, an independent control of the output current is possible via the piezoresistive transconductance. However, the size of the mentioned resonators is still in the order of micrometers and much more aggressive scaling can be achieved by using nanowire and nanotube structures.

Various top-down and bottom-up approaches have been proposed to create true NEMS structures, but very few of them have achieved ultrahigh resonance frequencies (>100 MHz). The top-down polysilicon NEM cantilever resonator

reported by Arcamone et al. [20] has the great merit of being fully integrated and interfaced with silicon CMOS circuitry; however, its resonance frequency is limited to 1.5 MHz with a quality factor of around 8,000 at applied bias of 2 V, in vacuum (0.9 Pa).

One of the first reports on very-high-frequency (VHF) NEM resonators describes platinum nanowires (43 nm in diameter and 1.3 μm long, suspended mass of 40 fg) resonating at frequencies higher than 100 MHz and quality factors of 8,500 at 4 K [50]. The same group (Roukes at Caltech) has reported later [51, 52] VHF NEM resonators based upon single-crystal nanowires (SiNWs) prepared by bottom-up synthesis. The metalized SiNW have resonance frequencies near 200 MHz and quality factors of the order of 2,000–2,500 (in vacuum at T =25 K) while the pristine nanowires have much higher quality factors (∼5,750 at 215 MHz). For these suspended nanowires, there is no coupling electrodes, the transduction is magnetomotive via the Lorentz force, $F(\omega) = L \cdot B \cdot I(\omega)$, generated by passing an RF current, $I(\omega)$, of variable frequency through the silicon nanowire. The Lorentz force excites the resonance in the suspended SiNWs, which induces an electromotive force and signal in the wires placed in a magnetic field B. Figure 3.18a,b report the scanning electron micrograph of a grown SiNW and the measured resonance characteristics of a metallized SiNW at VHF [51], with a fundamental resonance frequency at ∼200 MHz. The main reason invoked for metallizing the nanowires in these experiments (and obtaining impedance close to 50 Ω) is the impedance mismatch in pristine SiNW, which compromise the efficiency of actuation and transduction. In [52] the characteristics of SiNW resonators at room temperature with piezoelectric and electrostatic transductions are reported. The more practical electrostatic measurement schematic is shown in Fig. 3.18c and the corresponding signal characteristic at resonance (f_0 = 96 MHz, $Q \approx 550$) of a 40 −nm-thick, 1.8 −μm-long SiNW resonator, are given in Fig. 3.18d.

A material that attracted major interest for building NEM resonators, due to its high stiffness (E near 1 TPa), low density, defect-free structure, and ultrasmall cross section, is the carbon nanotube. Sazonova et al. [53] reported measurements of the resonant responses varying from 3 to 200 MHz with voltage-tunable characteristics, in CNTs with diameters of 1–4 nm, suspended over a trench. CNT NEM resonators with a fundamental mode resonance frequency higher than 1.3 GHz operating in ambient air have been reported by Peng et al. [54]. The device was created from single-wall (SW) CNT grown by chemical vapor deposition across a trench between metal source and drain electrodes on a silicon substrate (Fig. 3.19a). The method used to drive and investigate this resonator is depicted by the schematics in Fig. 3.19b. The gate is biased by a superposition of a DC signal, V_g, and an RF signal of frequency ω and amplitude δV_g, while on the drain a signal of frequency $2\omega - \Delta\omega$ and amplitude δV_d are used. The current is monitored at $\Delta\omega$ frequency by a lock-in technique; Fig. 3.19c shows data in vacuum and air with sharp changes in amplitude and phase at resonance (1.33 GHz). This CNT resonator can serve as a mass-sensitive sensor with exceptional sensitivity; when evaporating a very small amount of Fe on it, the authors were able to detect the mass with a resolution as small as 10^{-18}g.

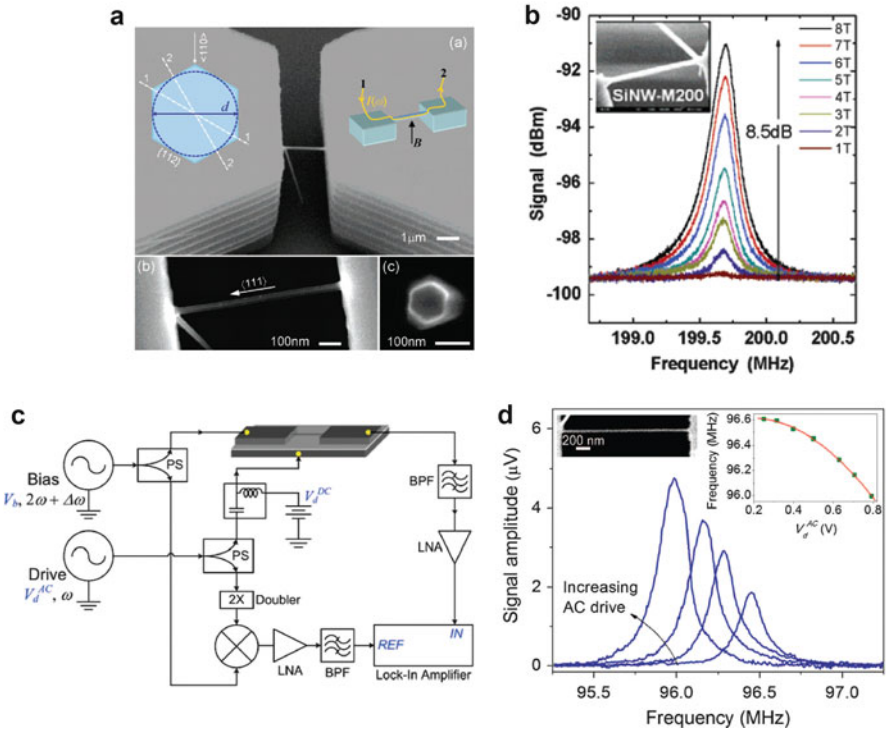

Fig. 3.18 (**a**) Bottom-up SiNW resonator excited in magnetic field and (**b**) measured resonance near 200 MHz. Reprinted with permission from [51]. ©2007 American Chemical Society. (**c**) Measurement setup for electrostatic transduction and (**d**) measured resonance signal in SiNW at room temperature. Reprinted with permission from [52]. ©2008 American Chemical Society

The NEM resonator with the highest resonance frequency reported to date is a similar CNT device from [54] loaded in an abacus style with inertial metal clamps (2.5 −nm-thick indium), yielding very short effective beam lengths [55]. When driven mechanically, such structure shows flexural modes in the very short segments between two adjacent indium beads, as described by the authors, this operation is akin to a musician placing the fingers (nanobeads) on the musical string (vibrating CNT) to select a specific musical note.

One of the major issues of such small vibrating SiNWs and CNTs is the early onset, a very low applied power, of nonlinearities characterized by frequency bistability arising from the effect of tension builtup in the wire at large vibration amplitudes, see Fig. 3.20a. A more general discussion concerning nonlinearities in MEM resonators, including electrical and mechanical nonlinearities was proposed in [56]. In fact, in NEM resonators, the linear dynamic range, bounded by the thermomechanical noise floor and the onset of nonlinearity, can practically disappear, meaning that nanowire and nanotube resonators are practically nonlinear

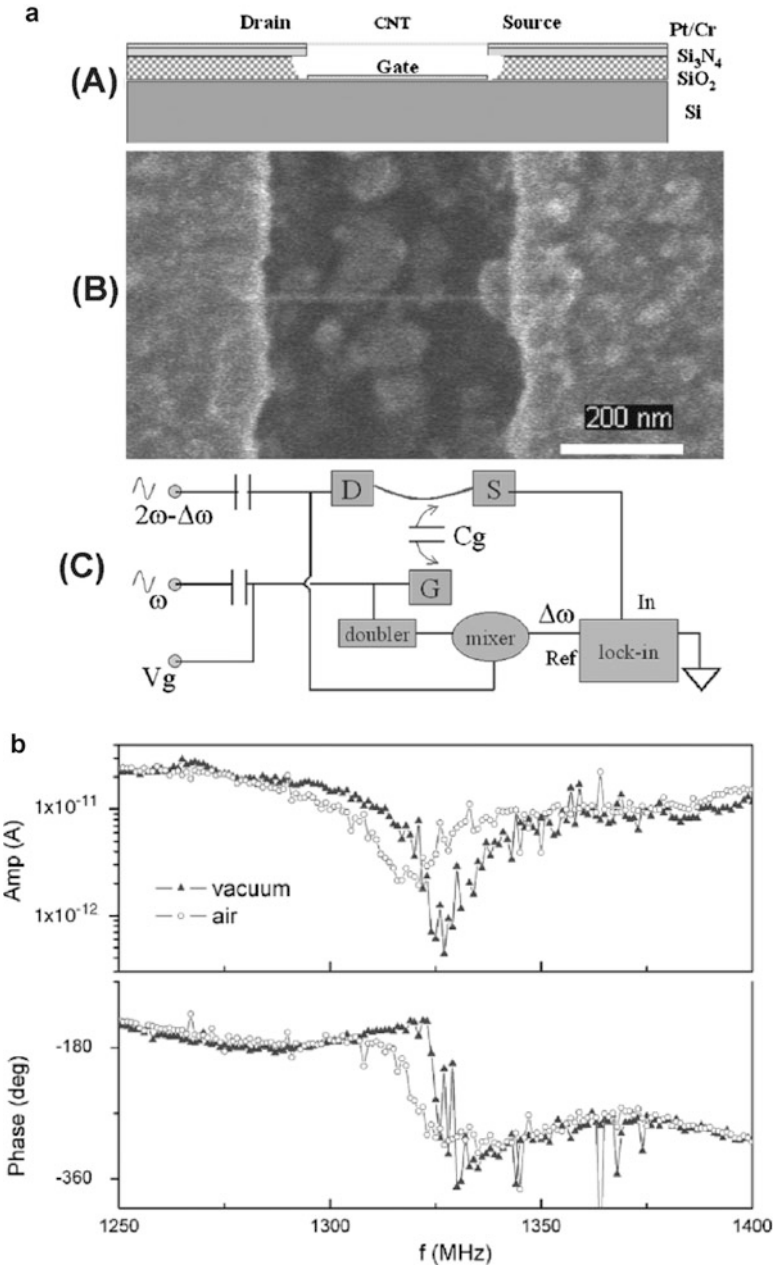

Fig. 3.19 (**a**) Schematic cross section of the CNT resonator. (**b**) SEM image of a suspended CVD-grown carbon nanotube crossing a trench. Scale bar is 200 nm. (**c**) Experimental diagram for the two mixer methods. Reprinted with permission from [54], ©2006 by American Physical Society

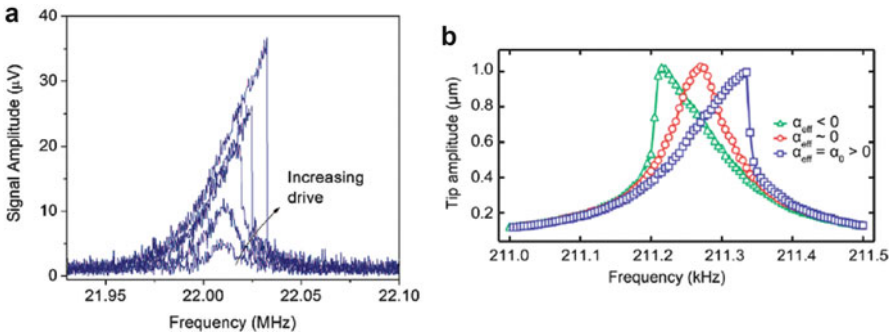

Fig. 3.20 (**a**) Nonlinearity at large vibration amplitudes observed in SiNW resonator. Reprinted with permission from [52]. ©2008 American Chemical Society. (**b**) Frequency response curves of the SiNW with different values of α_{eff}; the native nonlinearity is $\alpha_0 = 1.09 \times 10^{21}\ \mathrm{m}^{-2}\,\mathrm{s}^{-2}$. Reprinted with permission from [57]. ©2009 American Institute of Physics

elements [52]. The dominant nonlinearity in many NEMS is the cubic restoring force, often called the Duffing nonlinearity. Nichol et al. [57] have suggested that a nonlinear feedback can be used to cancel out the native nonlinearity or even change its sign (coefficient α_{eff} in Fig. 3.20b). The equations governing the nonlinearity in clamped–clamped nanowire and nanotube structures have been derived by Postma et al. [58]; they reported a strong dependence of the dynamic range on the structure aspect ratio. For some applications like noise squeezing, signal amplification, and mechanical mixing, the nonlinearities can be smartly exploited. However, for some other applications requiring a large linear dynamic range, such as ultrasensitive mass detection, or for building reference oscillators, they can be detrimental.

One solution proposed to limit the power applied on a single vibrating structure and reduce the overall motional resistance is to operate the M/NEM resonators in array configurations [59]. Therefore, the maximum power handling for a mechanically coupled array of resonators such as the ones shown in Fig. 3.21 is expressed as:

$$P_{\max n} = \omega_0 \frac{k_{eff\,n}}{Q_n} (a d_0)^2 = \omega_0 \frac{n \cdot k_{eff}}{Q_n} (a d_0)^2, \tag{3.8}$$

where n is the number of resonators in the array, and $k_{eff\,n}$ and Q_n are the stiffness and quality factor of the resonator array, respectively. The motional resistance is reduced (output signal level increased) and the phase noise is significantly improved (down by 17–26 dB in [59]). Figure 3.21b proves the significant improvement in terms of power handling, with a moderate degradation of the quality factor, when an array configuration is used. It is worth noting that such array technique has been successfully reported only for beams with micrometer sizes; it is probably much more difficult to apply it to vibrating nanowires and nanotubes due to the need of extremely small coupling elements.

Recently, the graphene material attracted further attention for its outstanding carrier mobility offered for high-ion transistors. The graphite material consists

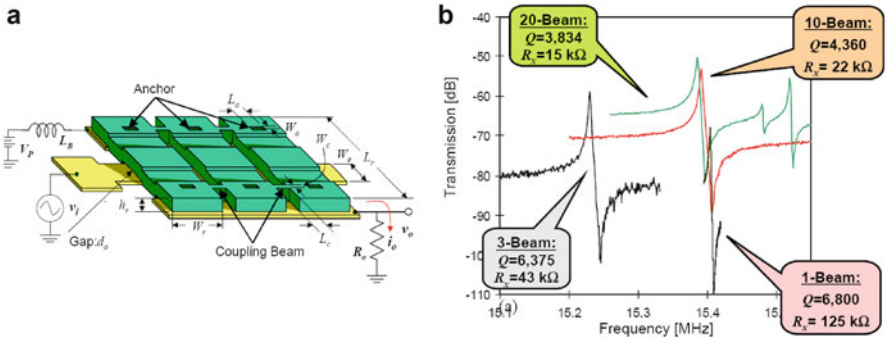

Fig. 3.21 (**a**) Depiction of arrays of parallel connected resonators in multibeam configuration to achieve higher level signal handling with smaller motional resistances. (b) Measured transmission characteristics; quality factor, Q; and motional resistance, R_x, for the various fabricated arrays with 1-, 3-, 10-, and 20-parallel-beam configuration. ©2004 IEEE. Reprinted, with permission, from [59]

of stacked layers of graphene sheets separated by 0.3 nm and held together by weak van der Waals forces, having extremely high strength, stiffness, and thermal conductivity along the basal plane. In [60] graphite is exfoliated onto an insulating substrate, producing micron-sized graphene sheets with thicknesses down to a single atomic layer that are suspended to form two-dimensional NEMS with ultimate thickness. Figure 3.22a depicts a resonant graphene sheet made by this technique with a thickness of 0.9 nm. The resulting resonators are excited both optically and electrically and produce, in vacuum conditions, resonance frequencies from 1 to 170 MHz (depending on the membrane thickness), with quality factors Q of 20–850. A graphene resonator with $f_0 = 70.5$ MHz, for the fundamental mode of the single layer, is shown in Fig. 3.22b. In this ultrathin vibrating NEMS, the role of the tension existing in the atomic-thin sheet is found to be of major importance for calculating the values of the resonance frequency:

$$f_0 = \sqrt{\left(A\sqrt{\frac{E}{\rho}}\,\frac{t}{L^2}\right)^2 + A^2\,0.57\,\frac{T}{\rho}\,L^2\,w\,t}, \tag{3.9}$$

where E is the Young modulus; ρ is the mass density; t, w, and L are the dimensions of the suspended graphene sheet; and the clamping coefficient, A, is 1.03 for doubly clamped beams. The built-in tension is $T = 13$ nN [60].

Beyond application in sensing (mass, force, and charge), the robust graphene ultrathin conductive membranes can act as a nanoscale supporting structure or atomically thin layer separating disparate environments.

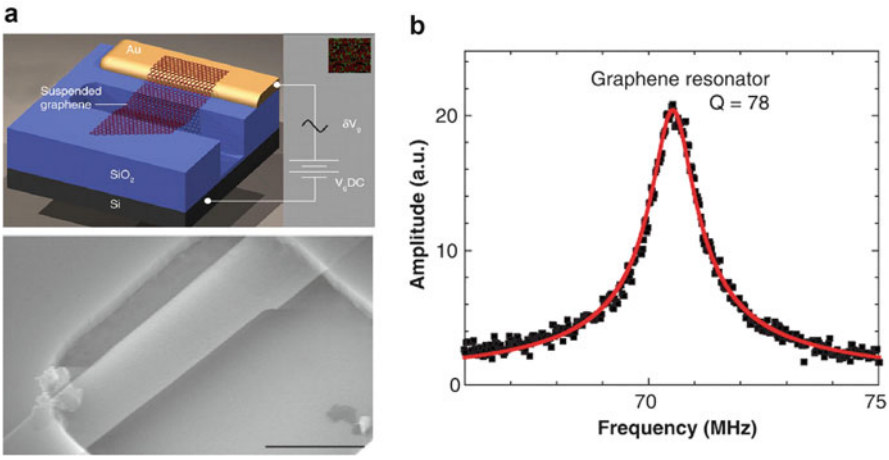

Fig. 3.22 (**a**) *Top:* schematic of a suspended graphene resonator and *Bottom:* image of a few-layer (∼2) graphene resonator; the scale bar is 1 μm. (**b**) Amplitude versus frequency taken with optical drive for the fundamental mode of the graphene resonator. From [60]. Reprinted with permission from AAAS

3.6.3 From Resonant Gate to Vibrating Body Transistors

The majority of the proposed and investigated MEM resonators exploit the principles of capacitive excitation and detection via very narrow air gaps. The main reason is the simplicity of the operation, the low power consumption, and a certain straightforwardness to design and fabricate such resonant structures made on semiconductors (monocrystalline silicon or polysilicon). In general, the capacitively transduced signals are very small and when the extreme scaling and the impedance matching of the resonators are also considered, some practical limitations could arise. Alternatively, the movable gate and body FET transistor structures proposed in Fig. 3.10a–d can operate as M/NEM resonators, with the main difference that the output is the drain current of the transistor, offering the possibility of building active resonators, similar to the very first proposed resonant gate transistor [31].

The first resonant gate transistors with modern silicon technology have been reported by EPFL; first, with out-of-plane AlSi resonant gate MOSFETs by Abelé et al. [61,62] and second, with in-plane resonant silicon gate transistors with various geometries by D. Grogg et al. [63]. The in-plane resonant gate MOSFET is an all-silicon resonator device suited to front-end integration, with the drain and source placed in the same horizontal plane as the gate (all made of the silicon layer of an SOI substrate), see Fig. 3.23a. Aggressively scaled versions of the in-plane resonant gate transistors have been reported by Durand et al. at LETI-CEA [64,65] based on silicon-on-nothing technology (with SiGe sacrificial layer) to achieve sub-100 −nm gaps and 400 −nm-thick single crystal resonators with a front-end process. A typical fabricated lateral resonant gate device with the length of 4 μm and the width of

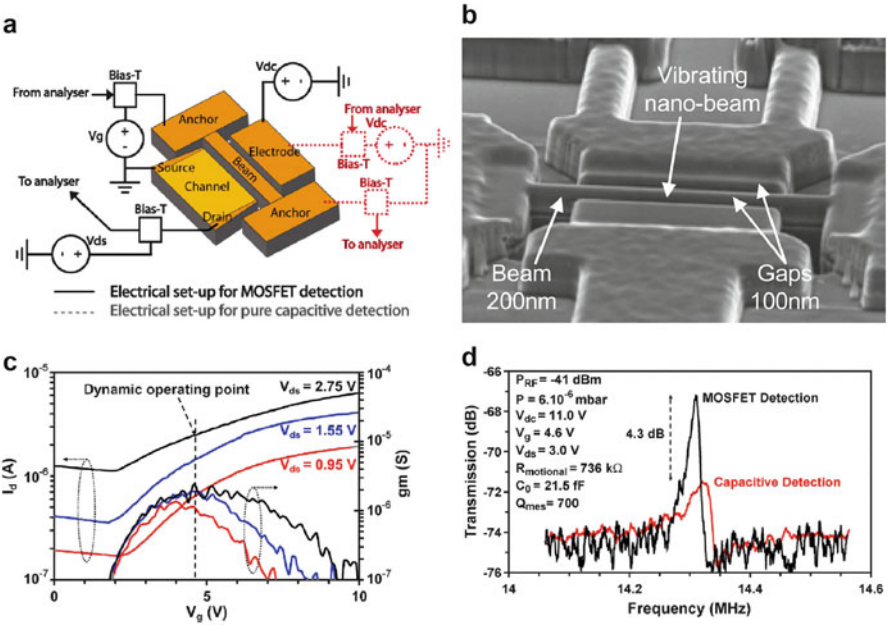

Fig. 3.23 (**a**) Schematic and measurement setup for resonant gate (RG-)MOSFET, (**b**) SEM image of fabricated RG-MOSFET, ©2008 IEEE. Reprinted, with permission, from [64], (**c**) static characteristics of the device, ©2009 IEEE. Reprinted, with permission, from [65], (**d**) corresponding measured transmission under vacuum showing resonance and +4.3-dB gain in RG-MOSFET configuration compared to capacitive detection

200 nm is depicted in Fig. 3.23b. The lateral MOS transistor could suffer from poor carrier mobility due to the roughness of the vertically etched sidewalls and have shown very little gain (the transistor gain being limited because of the very narrow width, defined by the thickness of the silicon film). Moreover, in [64], the channel surface was partially implanted on a few nanometers, resulting in a $2.1-\mathrm{M}\Omega$ resistor in parallel with the main transistor, explaining the relatively high leakage current (see Fig. 3.25c). However, the measured resonator transmission at an applied DC gate voltage maximizing the transconductance (e.g., gain) shows a resonance at 14.3 MHz and demonstrates a clear gain of +4.3 dB when the MOS detection is used instead of traditional capacitive detection. The unique ability of lateral resonant gate transistors to be integrated with advanced CMOS was demonstrated by Colinet et al. in [65], where a 130 −nm CMOS ASIC using a bridge measurement technique and a high sensitive first stage to minimize the influence of parasitic capacitances is monolithically combined with a resonant gate transistor to detect nanometer-scale displacements.

D. Grog et al. [66, 67] have proposed an alternative resonant transistor, called vibrating body FET (VB-FET). This device integrates a MOSFET transistor in a double-gate FIN-FET like suspended architecture with two side channels, and two

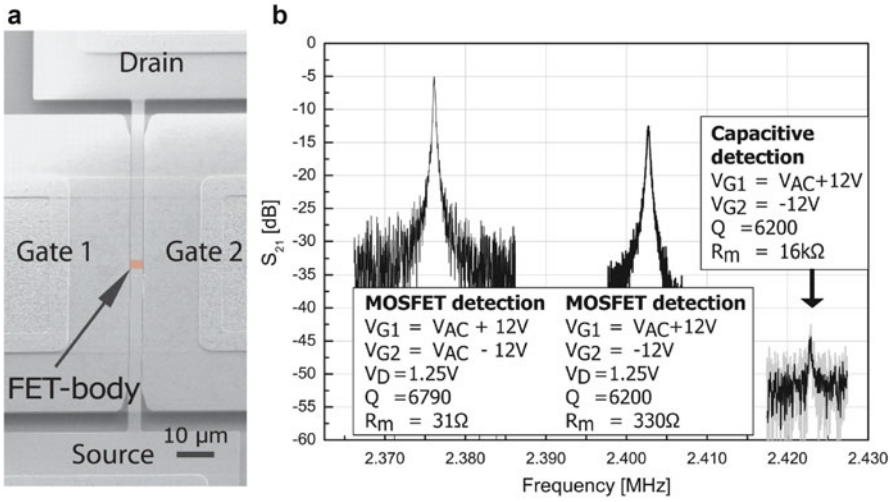

Fig. 3.24 (**a**) SEM image of a fabricated vibrating body field effect transistor (VB-FET) and (**b**) comparison of MOSFET detection for two different bias voltages with capacitive detection using the same structure. A gain of more than +30 dB is obtained with the MOSFET detection. ©2008 IEEE. Reprinted, with permission, from [67]

lateral gates separated from the transistor body by two air gaps, see Figs. 3.24a and 3.25a. In the VB-FET, the movable body modulates both the inversion or accumulation charge in the lateral channels and the piezoresistance of the structure (carrier mobility and mass). The piezoresistance modulation could be an extremely important mechanism in vibrating wires with very small cross sections and with low doping levels. SiNW show an unusually large piezoresistance effect compared with bulk. For example, the longitudinal piezoresistance coefficient along the $<111>$ direction increases with decreasing the diameter for p-type Si nanowires, reaching as high as $-3,550 \times 10^{-11}\,\mathrm{Pa}^{-1}$, in comparison with a bulk value of $-94 \times 10^{-11}\,\mathrm{Pa}^{-1}$ [68]. This giant piezoresistance effect can be exploited in future nanowire implementations of VB-FETs.

An outstanding gain of more than +30 dB for the output signal is obtained in micrometer-scale double-gate VB-FET when the output is taken on the transistor drain, compared with the identical structure operated in capacitive detection mode, Fig. 3.24b. Moreover, the device motional resistance is reduced from 16 kΩ below 100 Ω, which enables excellent conditions for 50 $-\Omega$ matching in RF applications. Figure 3.25a, b depict the depletion-mode double-gate structure reported in [64, 65] and its equivalent small-signal circuit with two gain mechanism: (a) channel charge modulation, for example, $i_{dFET} = g_{mFET} v_{in}$, and (b) piezoresistive, for example, $i_{piezo} = g_{mpiezo} v_{in}$. The transistor gain can be increased by increasing its transconductance, g_{mFET}, by applying higher gate and drain voltages up to eventually obtaining an active electromechanical resonator, with intrinsic gain. This unique property is demonstrated by the experimental data reported in Fig. 3.25,

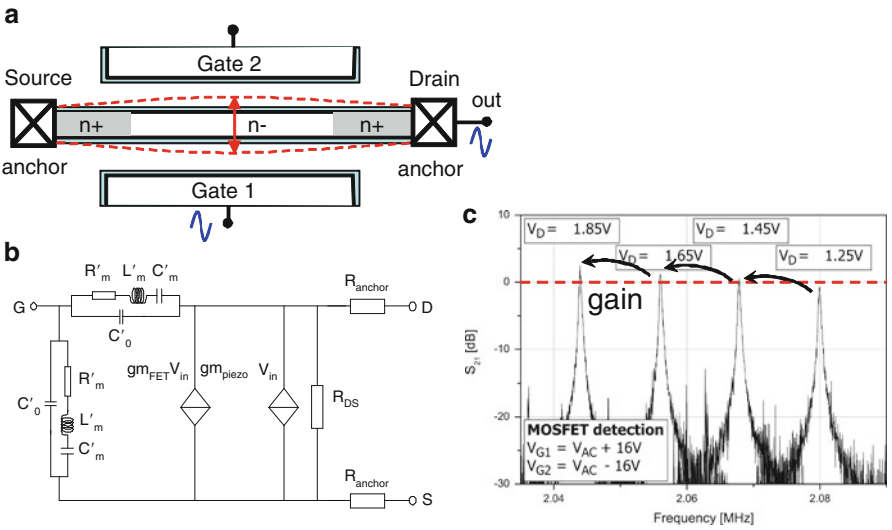

Fig. 3.25 (**a**) Depiction of double-gate in-plane vibrating body FET, (**b**) equivalent small-signal circuit of VB-FET showing the two main gain mechanisms: charge and piezoresistive, and (**c**) experimental s_{21} characteristics of a clamped–clamped beam VB-FET demonstrating the possibility to control and increase the gain with the applied drain voltage bias. An intrinsic gain >0 dB can be obtained with the VB-FET, suggesting the possibility to minimize or even to eliminate the sustaining amplifier. ©2009 IEEE. Reprinted, with permission, from [69]

where for $V_{ds} > 1.25$ V the resonance peak reaches a level superior to 0 dB, for instance, $+3$ dB at $V_{ds} = 1.65$ V (in reality the gain is even higher, the measurement being carried out with an imperfect impedance matching). This result suggests that the gain mechanism in VB-FET can be exploited to simplify the design of the sustaining oscillator and obtain ultralow-power oscillators with sub-10 $-\mu$W power consumption [69,70] or, even further, provide self-sustained oscillations without the need of any external sustaining amplifier (e.g., a single-device oscillator) [71].

While the clamped–clamped beam design of the VB-FET is the most straightforward implementation for flexural-mode resonators, Fig. 3.26 shows that multigate VB-FET resonators operating at much higher frequency in bulk mode can be achieved. In this particular design, a lateral transistor and corresponding gate are defined on each side of a square resonator. The transistor source contacts are connected together and one can activate one, two, or four gates to take advantage of the VB-FET operation compared to capacitive operation (Fig. 3.26b). Interestingly, in the square VB-FET, vibrating in a bulk extensional mode, the piezoresistive modulation is found to provide an important contribution to the output signal [69, 70].

Another active resonator has been proposed by van Beek et al. [72] by using the mechanical strain rather than an electrical field to modulate the conductivity of the silicon. The resonator layout is depicted in Fig. 3.27a and can be represented

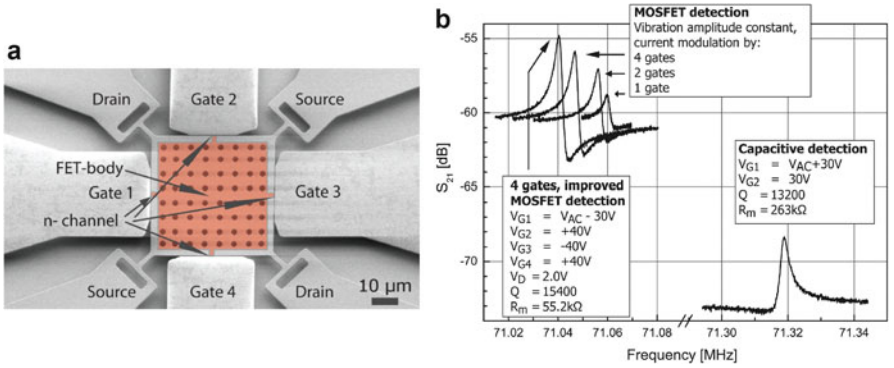

Fig. 3.26 (**a**) Multigate VB-FET device in a 4-gate implementation on SOI and (**b**) measured characteristics of a square 4-gate multigate VB-FET as a function of the number of active gates, compared to capacitive detection using a single gate. ©2008 IEEE. Reprinted, with permission, from [67]

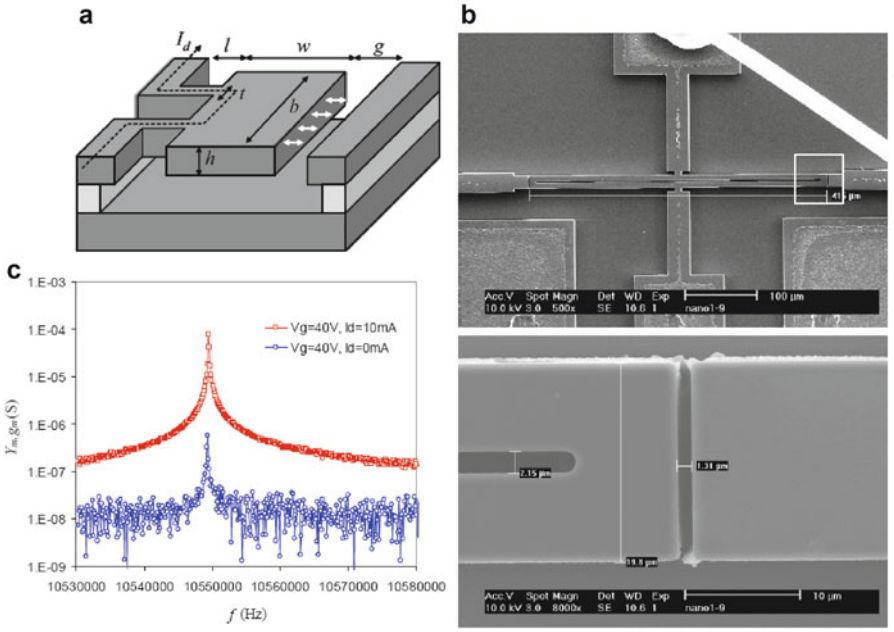

Fig. 3.27 (**a**) Piezoresistive resonator principle and demonstrated gain compared with capacitive operation of the same structure (**b**) SEM fabricated device. ©2008 IEEE. Reprinted, with permission, from [72]

by a lumped mass-spring system. The springs of the system are approximately localized in the "arms" of the resonator with length l, width t, and height h. The mass is localized in the resonator's "head" with length w, width b, and height h.

At resonance, the resonator arms are compressed and stretched along their axis. Therefore, the resulting mechanical strain is detected by the change in electrical resistance of the resonator arms due to the piezoresistive effect. The authors have implemented this resonator in 1.5 −μm-thick SOI and proposed a small-signal equivalent model, with the gain calculated according to the expression:

$$\frac{v_d}{v_g} = \frac{g_m(\omega)}{g_d} = jQ\varepsilon_0\pi_i V_{g0} V_{d0} \frac{b}{2tg^2}, \tag{3.10}$$

where the piezoresistive coefficient π_i is maximum when the resonator arm is oriented in the <100> direction. This resonator can provide a gain higher than unity, similarly to the VB-FET. Figure 3.27c shows experimental results with an increased signal when the device is operated as a transistor. The gain can be improved by decreasing the gap width, g, or by increasing the ratio of the arm width to the head width, b/t. Amplification factors up to 4.6 dB and Q values up to 60,000 were obtained for 15 MHz resonators. This piezotransistor principle is expected to be particularly suited for future silicon nanowire resonators.

3.7 NEM Mixers and Single Nanotube Radio

In communication applications, the mixer function is extremely useful for any transceiver architecture. Early mixer operation of MEM resonators has been reported by Nguyen [73] by applying radio frequency ($v_{RF} = V_{RF} \cdot \cos \omega t$) and low-frequency ($v_{LO} = V_{LO} \cdot \cos \omega t$) signal to the two electrodes of a vibrating beam in Fig. 3.13a and by exploiting the nonlinear voltage-to-force transduction of the resonator, relating the input force to input voltage ($v_{RF} - v_{LO}$) by a square law:

$$F_d = \ldots + \frac{1}{2} V_{RF} V_{LO} \frac{\partial C}{\partial x} \cos(\omega_{RF} - \omega_{LO})t + \ldots. \tag{3.11}$$

The operation schematic and the spectrum of a single-device vibrating body FET mixer-filter, where the RF and the LO components to be mixed are applied on the same electrode, is depicted in Fig. 3.28. The resonator behaves itself as a filter, providing a relevant output signal only if the difference of the two applied frequencies matches the mechanical resonance frequency (for the VB-FET mixer from [67], $f_0 = 9.84$ MHz). MEM mixers can provide major simplification of RF front-ends with the simultaneous reduction of the power consumption; for instance, the power consumption of the active VB-FET mixer reported in [67] is only 500 μW.

A more sophisticated exploitation of RF NEMS for receivers is illustrated by the all-in-one nanotube radio, proposed by Jensed et al. [74]. A vibrating carbon nanotube serves simultaneously as all essential components of a radio: antenna, tunable bandpass filter, amplifier, and demodulator (Fig. 3.29). A direct current (dc) voltage source is connected to the CNT electrodes and powers the radio. The applied

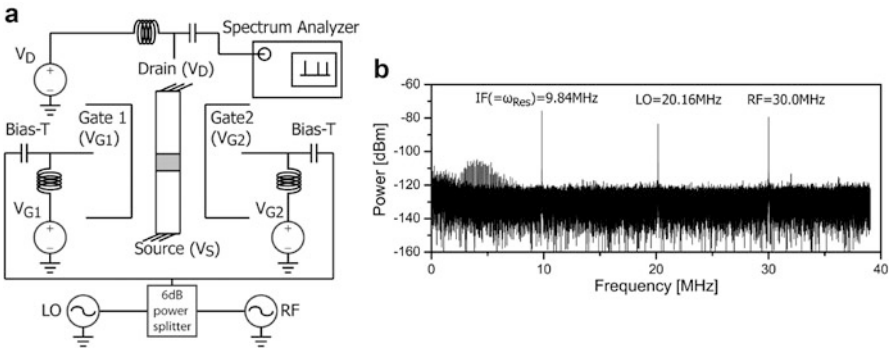

Fig. 3.28 (**a**) VB-FET mixer configuration and (**b**) experimental demonstration of the mixer-filter with $f_{high} - f_{low} = f_{res} = 9.84$ MHz. ©2008 IEEE. Reprinted, with permission, from [67]

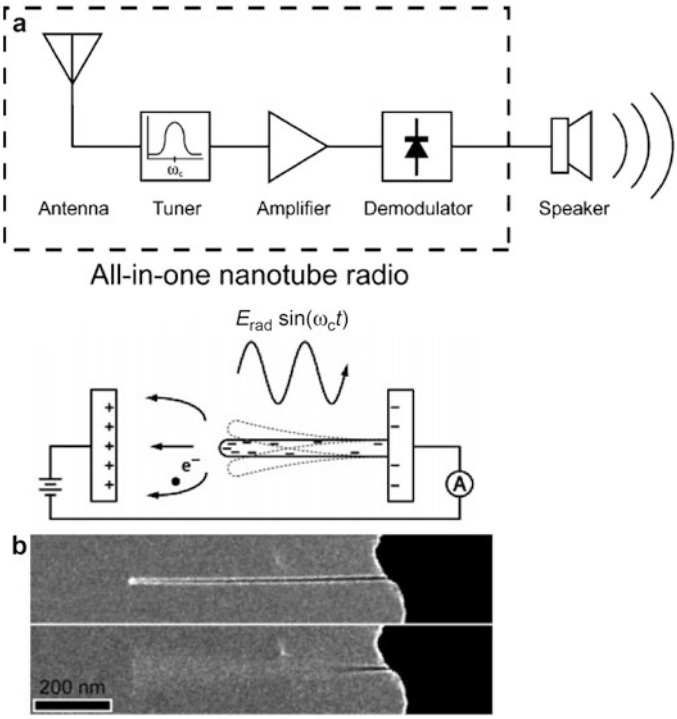

Fig. 3.29 (**a**) Principle of all-in-one nanotube radio, (**b**) fabricated device used to experimentally proof the concept. Reprinted with permission from [74]. ©2007 American Chemical Society

dc bias negatively charges the tip of the nanotube, placed in advanced vacuum, sensitizing it to oscillating electric fields. The principle of the nanotube radio is quite different from the one used in traditional radios: the electromagnetic waves from

a radio transmission impinge upon the nanotube, forcing it to physically vibrate through their action on the charged tip. Of great importance is that the vibrations are only significant when the frequency of the incoming wave coincides with the CNT flexural resonance frequency. Successful reception has been demonstrated with carrier waves in the 40–400 −MHz range and both frequency and amplitude modulation techniques. This demonstration of a completely novel functionality embedded in a single NEM resonator opens the door to many other future low-power communication applications.

3.8 Conclusions and Perspectives

The principles and examples showed in this chapter support the idea that NEMS-based information processing offers interesting opportunities for future digital and analog/RF applications. The novel functionality and the low-power consumption of NEM devices are the two main opportunities to seize. Moreover, NEMS processing (top-down or bottom-up) is compatible with advanced silicon ICs, which can result in high-performance, low-power hybrid NEMS-CMOS systems. Particularly the small mass and high stiffness of silicon- or carbon-based NEM resonators offers unique features for high-frequency (>100 MHz–1 GHz) applications such as communication circuits. The resulting miniaturization and the low costs are other key benefits. Beyond analog and digital applications, NEMS are also foreseen as unique candidates for pushing the limits of sensing (mass, force, gas, and biosensing) due to their extreme sensitivities to the mass loading or to the very small amount of electrical charges. This important class of sensing applications was not the object of this chapter.

Additionally, it is important to note that future design of RF ICs with NEMS should not be limited to mimicking the traditional circuit architectures and having the NEMS just as replacement components for some parts of the integrated circuits or for off-chip components. The promise of NEMS consists in being able to embed full circuit functions and generate novel circuit design, more compact, with more functionality and lower power consumption. Certainly, there are still a lot of problems to solve before we will see NEMS devices in mass market applications like the reliability and packaging that have not been described in the chapter due to limited available space. However, all the recent progress achieved in terms of packaging and reliability for MEMS is exploitable and applicable to NEMS. Of course, there are some NEMS-specific phenomena like the van der Waals and Casimir forces that should be accounted for when one deals with more aggressive scaling and optimized design, but globally, many of the processes and design advancement in MEMS field can be used and exploited in NEMS, similar to the way the various progresses in CMOS technology have been used as an additive booster strategy to improve its performance at nanoscale.

Finally, in this chapter we have presented some more disruptive concepts such as the hybrid NEM-solid-state devices, which could embed a transistor in a movable

object. Such device architecture can take benefits from both the novel NEM functionality and the signal gain that is specific to any transistor. In the case of vibrating transistors, this approach results in a novel category of devices that can be called active NEM resonators, exploiting charge and/or piezoresistive modulations. The active NEM resonator can further simplify the IC design and offer more complex functions in a single device; this is certainly one of the most interesting way to pursue in the future for the success of NEMS.

References

1. C.T.C. Nguyen, IEEE Trans. Microw. Theor. Tech. **47**(8), 1486 (1999)
2. C.T.C. Nguyen, IEEE Trans. Ultrason. Ferroelectrics Freq. Contr. **54**(2), 251 (2007)
3. M.L. Roukes, in *Digest of IEEE International Electron Devices Meeting (IEDM)*, 2004, pp. 539–542
4. M.L. Roukes, in *IEEE 20th Interational Conference on Micro Electro Mechanical Systems (MEMS 2007)*, 2007, pp. 93–94
5. J.M. Bustillo, R.T. Howe, R.S. Muller, Proc. IEEE **86**(8), 1552 (1998)
6. Z. Cui, C. Gu, in *1st IEEE Interational Conference on Nano/Micro Engineered and Molecular Systems (NEMS '06)*, 2006, pp. 607–610
7. E. Ollier, L. Duraffourg, M.T. Delaye, H. Grange, S. Deneuville, J. Bernos, R. Dianoux, F. Marchi, D. Renaud, T. Baron, P. Andreucci, P. Robert, in *2nd IEEE Int. Conf. on Nano/Micro Engineered and Molecular Systems (NEMS '07)*, 2007, pp. 180–185
8. S.Y. Chou, P.R. Krauss, P.J. Renstrom, J. Vac. Sci. Technol. B **14**(6), 4129 (1996)
9. J. Arcamone, M.A.F. v. d. Boogaart, F. Serra-Graells, J. Fraxedas, J. Brugger, F. Perez-Murano, Nanotechnology **19**, 302 (2008)
10. Y.K. Choi, T.J. King, C. Hu, in *2001 Interational Semiconductor Device Research Symposium*, 2001, pp. 543–546
11. Y.J. Chen, J.H. Hsu, H.N. Lin, Nanotechnology **16**, 1112 (2005)
12. F. Laermer, A. Urban, in *Digest of the 13*th *Int. Conf. on Solid-State Sensors, Actuators and Microsystems (Transducers 2005)*, vol. 2, 2005, pp. 1118–1121
13. J. Lianga, F. Kohsakaa, T. Matsuoa, X. Lia, T.S. Uedaa, Microelectronic Engineering **85**(5-6), 1000 (2008)
14. V. Pott, A.M. Ionescu, Microelectronic Engineering **83**(4-9), 1718 (2006)
15. J. Appenzeller, J. Knoch, M.T. Bjork, H. Riel, H. Schmid, W. Riess, IEEE Trans. Electron Dev. **55**(11), 2827 (2008)
16. S. Iijima, Nature **354**, 56 (1991)
17. P. Zhang, Y. Huang, P.H. Geubelle, P.A. Klein, K.C. Hwang, Int. J. Solid Struct. **39**, 3894 (2002)
18. B. Bhushan, *Handbook of Nanotechnology* (Springer, New York, 2004)
19. M. Li, R.B. Bhiladvala, T.J. Morrow, J.A. Sioss, K.K. Lew, J.M. Redwing, C.D. Keating, T.S. Mayer, Nat. Nanotechnol. **3**, 88 (2008)
20. J. Arcamone, B. Misischi, F. Serra-Graells, M.A.F. v. d. Boogaart, J. Brugger, F. Torres, G. Abadal, N. Barniol, F. Pérez-Murano, IEEE Trans. Circ. Syst. II **54**(5), 377 (2007)
21. C. Stampfer, A. Jungen, C. Hierold, IEEE Sensor J. **6**(3), 613 (2006)
22. D. Tsamados, A.M. Ionescu, K. Akarvardar, H.S.P. Wong, E. Alon, T.J.K. Liu, White paper ERD ITRS on Nano-Electro-Mechanical Switches. Techenical Report (2007)
23. K. Akarvardar, D. Elata, R. Parsa, G.C. Wan, K. Yoo, J. Provine, P. Peumans, R.T. Howe, H.S.P. Wong, in *Digest of IEEE Interational Electron Devices Meeting (IEDM)*, 2007, pp. 299–302

24. F. Chen, H. Kam, D. Markovic, T.J.K. Liu, V. Stojanovic, V.E. Alon, in *IEEE/ACM Interational Conference on Computer-Aided Design (ICCAD)*, 2008, pp. 750–757
25. R.S. Chakraborty, S. Narasimhan, S. Bhunia, IEEE Trans. Circ. Syst. I **54**(11), 2480 (2007)
26. Y. Zhou, S. Thekkel, S. Bhunia, in *Proceedings of the Interational Symposium on Low-Power Electronics Design (ISLPED)*, 2007, pp. 14–19
27. G.M. Rebeiz, *RF MEMS: Theory, Design and Technology* (Wiley, New York, 2003)
28. W.W. Jang, J.O. Lee, J.B. Yoon, M.S. Kim, J.M. Lee, S.M. Kim, K.H. Cho, D.W. Kim, D. Park, W.S. Lee, Appl. Phys. Lett. **92**(10), 103 (2008)
29. A.B. Kaul, E.W. Wong, L. Epp, B.D. Hunt, Nano Lett. **6**(5), 942 (2006)
30. Y. Hayamizu, T. Yamada, K. Mizuno, R.C. Davis, D.N. Futaba, M. Yumura, K. Hata, Nat. Nanotechnol. **3**, 289 (2008)
31. H.C. Nathanson, W.E. Newell, R.A. Wickstrom, J.R. Davis, IEEE Trans. Electron Dev. **14**(3), 117 (1967)
32. I. Eisele, M. Zimmer, in *Digest of Device Research Conference (DRC)*, 2002, pp. 113–116
33. A.M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. Declercq, P. Renaud, C. Hibert, P. Fluckiger, G.A. Racine, in *Proceedings of the International Symposium on Quality Electronic Design*, 2002, pp. 496–501
34. N. Abele, A. Villaret, A. Gangadharaiah, C. Gabioud, P. Ancey, A. Ionescu, in *Electron Devices Meeting, 2006. IEDM '06. International*, 2006, pp. 1–4. DOI 10.1109/IEDM.2006.346826
35. H. Kam, D. Lee, R. Howe, T.J. King, in *Digest of IEEE Interational Electron Devices Meeting (IEDM)*, 2005, pp. 463–466
36. D. Grogg, C. Meinen, D. Tsamados, H.C. Tekin, M. Kayal, A.M. Ionescu, in *Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, 2008, pp. 302–305
37. B. Pruvost, H. Mizuta, S. Oda, IEEE Trans. Nanotechnol. **6**(2), 218 (2007)
38. K. Akarvardar, C. Eggimann, D. Tsamados, Y. Chauhan, G.C. Wan, A.M. Ionescu, R.T. Howe, H.S.P. Wong, IEEE Trans. Electron Dev. **55**(1), 48 (2008)
39. Y.S. Chauhan, D. Tsamados, N. Abele, C. Eggimann, M. Declercq, A.M. Ionescu, in *IEEE Conference on VLSI Design*, 2008
40. E. Colinet, C. Durand, P. Audebert, P. Renaux, D. Mercier, L. Duraffourg, E. Ollier, F. Casset, P. Ancey, L. Buchaillot, A.M. Ionescu, in *Digest of the IEEE Interational Conference on Solid-State Circuits (ISSCC)*, 2008, pp. 332–333
41. C.T.C. Nguyen, IEEE Trans. Ultrason. Ferroelectrics Freq. Contr. **54**(2), 251 (2007)
42. M.L. Roukes, in *Proceedings of IEEE MEMS*, 2007, pp. 93–94
43. K. Jensen, H.B. Peng, A. Zetl, in *Proceedings of the Interational Conference on Nanoscience and Nanotechnology (ICONN)*, 2006, pp. 68–71
44. H. Lakdawala, X. Zhu, H. Luo, S. Santhanam, L.R. Carley, G.K. Fedder, IEEE J. Solid State Circ. **37**(3), 394 (2002)
45. J.R. Vig, Y. Kim, IEEE Trans. Ultrason. Ferroelectrics Freq. Contr. **46**(6), 1558 (1999)
46. J.R. Vig, C. Audion, L.S. Cutter, M.M. Driscoll, E.P. Eernisse, R.L. Filler, R.M. Garvey, W.J. Riley, R.C. Smythe, R.D. Weglein, in *Proceedings of IEEE Frequency Control Symposium*, 1992, pp. 763–781
47. Y. Xie, S.S. Li, Y.W. Lin, Z. Ren, C.T.C. Nguyen, in *Digest of IEEE Interational Electron Devices Meeting (IEDM)*, 2003, pp. 953–956
48. D. Weinstein, S. Bhave, in *Proceedings of Solid-State Sensors, Actuators and Microsystems Workshop*, 2008, pp. 368–371
49. D. Weinstein, S. Bhave, in *Digest of IEEE Interational Electron Devices Meeting (IEDM)*, 2007, pp. 415–418
50. A. Hussain, J. Hone, H.W. Postma, X.M.H. Huang, T. Drake, M. Narbic, A. Scherer, M.L. Roukes, Appl. Phys. Lett. **83**(6), 1240 (2003)
51. X.L. Feng, R. He, P. Yang, M.L. Roukes, Nano Lett. **7**(7), 1953 (2007)
52. R. He, X.L. Feng, M.L. Roukes, P. Yang, Nano Lett. **8**(6), 1756 (2008)
53. V. Sazonova, Y. Yaish, H. Ustunel, D. Roundy, T.A. Arlas, P.L. McEuen, Nature **431**(Sep.), 284 (2004)

54. H.B. Peng, C.W. Chang, S. Aloni, T.D. Yuzvinsky, A. Zettl, Phys. Rev. Lett. **PRL 97**, 087203 (2006)
55. H.B. Peng, C.W. Chang, S. Aloni, T.D. Yuzvinsky, A. Zettl, Phys. Rev. B **76**, 035405 (2007)
56. M. Agarwal, K.K. Park, R.N. Candler, B. Kim, M.A. Hopcroft, S.A. Chan-dorkar, C.M. Jha, R. Melamud, T.W. Kenny, B. Murmann, in *Proceedings of IEEE Interational Frequency Control Symposium*, 2006, pp. 209–212
57. J.M. Nichol, E.R. Hemesath, L.J. Lauhon, R. Budakian, Appl. Phys. Lett. **95**, 123116 (2009)
58. H.W.C. Postma, I. Kozinsky, A. Husain, M.L. Roukes, Appl. Phys. Lett. **86**, 223105 (2005)
59. S. Lee, C.T.C. Nguyen, in *Proceedings of IEEE Interational Frequency Control Symposium*, 2004, pp. 144–150
60. J.S. Bunch, A.M. v. d. Zande, S.S. Verbridge, I.W. Frank, D.M. Tanenbaum, J.M. Parpia, H.G. Craighead, P.L. McEuen, Science **315**(490), 490 (2007)
61. N. Abele, R. Fritschi, K. Boucart, F. Casset, P. Ancey, A.M. Ionescu, in *Digest of IEEE Interational Electron Devices Meeting (IEDM)*, 2005, pp. 479–481
62. N. Abele, K. Segueni, K. Boucart, F. Casset, B. Legrand, L. Buchaillot, P. Ancey, A.M. Ionescu, in *19th IEEE Interational Conference on Micro Electro Mechanical Systems (MEMS)*, 2006, pp. 882–885
63. D. Grogg, D. Tsamados, N.D. Badila, A.M. Ionescu, in *Proceedings of International Solid-State Sensors, Actuators and Microsystems Conference Transducers*, 2007, pp. 1709–1712
64. C. Durand, F. Casset, P. Renaux, N. Abele, B. Legrand, D. Renaud, E. Ollier, P. Ancey, A.M. Ionescu, L. Buchaillot, IEEE Electron Device Lett. **29**(5), 494 (2008)
65. E. Colinet, C. Durand, L. Duraffourg, P. Audebert, G. Dumas, F. Casset, E. Ollier, P. Ancey, J.F. Carpentier, L. Buchaillot, A.M. Ionescu, IEEE J. Solid State Circ. **44**(1), 247 (2009)
66. D. Grogg, H.C. Tekin, N.D. Badila-Ciressan, M. Mazza, D. Tsamados, A.M. Ionescu, in *Digest of Annual Device Research Conference (DRC)*, 2008, pp. 155–156
67. D. Grogg, M. Mazza, D. Tsamados, A.M. Ionescu, in *Digest of IEEE Interational Electron Devices Meeting (IEDM)*, 2008, pp. 1–4
68. R. He, P. Yang, Nat. Nanotechnol. pp. 42–46 (2006)
69. D. Grogg, S. Ayöz, D. Tsamados, A.M. Ionescu, in *Proceedings of the European Solid State Device Research Conference (ESSDERC)*, 2009, pp. 379–382
70. D. Grogg, A.M. Ionescu, in *International Symposium on VLSI Technology, Systems, and Applications (VLSI-TSA)*, 2009, pp. 129–130
71. D. Grogg, S. Ayöz, A.M. Ionescu, in *Digest of IEEE International Electron Devices Meeting (IEDM)*, pp. 1–4 (2009)
72. J.T.M.v. Beek, K.L. Phan, G.J.A.M. Verheijden, G.E.J. Koops, C. v. d. Avoort, J. v. Wingerden, D.E. Badaroglu, J.J.M. Bontemps, R. Puers, in *Digest of IEEE International Electron Devices Meeting (IEDM)*, 2008, pp. 1–4
73. C.T.C. Nguyen, in *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting*, 2000, pp. 142–149
74. K. Jensen, J. Weldon, H. Garcia, A. Zettl, Nano Lett. **7**(11), 3508 (2007)

Chapter 4
Future Trends in Acoustic RF MEMS Devices

Bertrand Dubus

Abstract Piezoelectricity and longitudinal elastic wave propagation constitute the basic physical mechanisms involved in the classical bulk acoustic wave resonator.[1] Innovative acoustic MEMS will rely on other types of elastic waves (shear waves, guided waves in free plates or plates bonded on substrate, waves in periodic media) or transduction mechanisms (electrostriction) which are described in the first section of this chapter. Operation and characteristics of emerging acoustic RF MEMS devices, such as shear and guided wave resonators, tunable resonators and phononic crystal-based resonators and filters, are reviewed in the other sections.

4.1 Overview of Physical Phenomena Used in Future RF MEMS Devices

In acoustic RF MEMS devices, electrical effects coupled to elastic[2] wave propagation are used to obtain a specific electrical response. Considering, for example, a classical bulk acoustic wave (BAW) resonator, the thickness resonance of the longitudinal wave in a piezoelectric layer having electrodes on top and bottom surfaces generates a rapid and large change of the input electrical impedance with

[1] Surface acoustic wave (SAW) devices are not discussed in this chapter except in Sect. 4.3. SAW devices are generally not considered as MEMS for historical (first SAW devices were developed when MEMS technology did not exist) and technological (classical SAW devices require a piezoelectric substrate) reasons. This separation is less and less justified today, as current research in SAW, BAW and other acoustic RF MEMS devices converges.

[2] In this chapter, the term "elastic waves" is preferred to "acoustic waves" which is also associated to waves propagating in fluids. However, devices are still denoted as "acoustic RF MEMS devices" throughout the text as they are often designated by these terms in electrical engineering.

B. Dubus (✉)
Institut d'Electronique de Microélectronique et de Nanotechnologie, Lille, France
e-mail: bertrand.dubus@isen.fr

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_4,

frequency. Such acoustic RF MEMS device can be divided schematically into three parts representing the mechanical part, the electromechanical transduction and the electrical part.

The function of an acoustic RF MEMS device is based on a physical effect associated to elastic wave propagation. Section 4.1.1 provides an overview on elastic wave propagation with a focus on specific properties which could be used in future RF MEMS devices. Analysis of elastic waves propagating in unbounded media, waveguides and periodic media is presented.

In the RF range, the electromechanical transduction is often obtained via piezoelectric effect. The efficiency to convert electrical energy into mechanical energy and vice versa is evaluated by a quantity called coupling coefficient. Many parameters affect coupling coefficient: material class of symmetry and physical constants, distribution of displacement and electric fields, geometry of the sample, position of the electrodes, etc. Section 4.1.2 gives a general description of the piezoelectric effect with a view to explain how changes of these parameters can lead to improved or new RF devices. Electrostriction which is another physical effect coupling electrical and mechanical fields in insulators is also discussed.

4.1.1 Elastic Wave Propagation

This section gives a brief description of elastic wave properties in solids. A detailed analysis can be found in the following textbooks: [1, 2] for bulk and guided waves, [3] for waves in periodic media.

4.1.1.1 Bulk Waves

In elastodynamics, the vibration of linear elastic materials is considered in a small strain approximation. From Newton's law and Hooke's law, equation of motion, not considering external forces, is written as

$$\rho \frac{\partial^2 u_i}{\partial t^2} = c_{ijkl} \frac{\partial^2 u_l}{\partial x_j \partial x_k}, \tag{4.1}$$

where u_i and u_l are the components of the displacement field, c_{ijkl} the elastic constants of the material, ρ the material density, x_j and x_k the space coordinates and t the time. Subscripts i to l vary from 1 to 3. Einstein's summation notation is used throughout the chapter. Equation (4.1) is reduced to an eigenvalue problem when plane waves are considered. In the case of an isotropic material, the elastic constants are expressed with two independent constants: Young's modulus E and Poisson's ratio ν. Three plane waves can propagate in any given direction **n**: one *longitudinal wave* (displacement along **n)** at velocity

$$V_L = \sqrt{E(1-\nu)/\rho(1+\nu)(1-2\nu)},$$

and two *transverse (or shear) waves* (displacement perpendicular to **n**) at velocity

$$V_T = \sqrt{E/2\rho(1+\nu)}.$$

In anisotropic materials, three plane waves can propagate in any given direction, one quasi-longitudinal wave and two quasi-transverse waves, each of these waves having its own velocity which depends upon propagation direction.

4.1.1.2 Guided Waves

Material boundaries have the effect of guiding elastic waves. The simplest waveguide is the free surface. In an elastic half-space, the surface or Rayleigh wave verifies (4.1) together with stress-free conditions at the surface. It is constituted by the superposition of longitudinal and transverse partial waves that cancel stresses at the surface. A Rayleigh wave propagates along the surface with displacement amplitude exponentially decreasing when moving perpendicular to the surface. It is non-dispersive but its polarization is not rectilinear.

An unbounded plate with horizontal stress-free surface constitutes another simple waveguide. In such geometry, the horizontally polarized transverse partial wave remains decoupled from longitudinal and transverse waves polarized in the vertical plane. At circular frequency ω, the guided waves constituted by the superposition of horizontally polarized transverse (or shear) partial waves that verify the stress-free conditions on both surfaces are denoted *SH waves*. The longitudinal and transverse partial waves polarized in the vertical plane undergo mode conversion for each reflection at the surfaces and are therefore coupled at the boundaries. The guided waves constituted by the superposition of longitudinal and transverse partial waves polarized in the vertical plane that verify the stress-free conditions on both surfaces are called *Lamb waves*.

In straight waveguides considered aligned along x_1 axis, displacement field components associated to a guided wave are expressed as

$$u_i(x_1,x_2,x_3,t) = \tilde{u}_i(x_2,x_3)\, e^{+j(\omega t-\beta x_1)}, \tag{4.2}$$

where β is the wavenumber. As in electromagnetism, properties of guided waves are analysed from dispersion curve representing ω (or frequency f) versus β. Figure 4.1 displays the dispersion curve of Lamb waves in a 1-µm-thick aluminium nitride (AlN) plate with some corresponding displacement fields. Several branches are found corresponding to different mode families. Lamb waves with symmetrical (respectively anti-symmetrical) displacement field with respect to plate mid-surface are denoted S (respectively A) waves. All these modes are dispersive as Lamb waves are constituted by the superposition of partial plane waves reflected on plate boundaries and therefore depend upon boundary separation distance. At $\beta = 0$, the longitudinal and transverse partial waves are decoupled, and branch points are

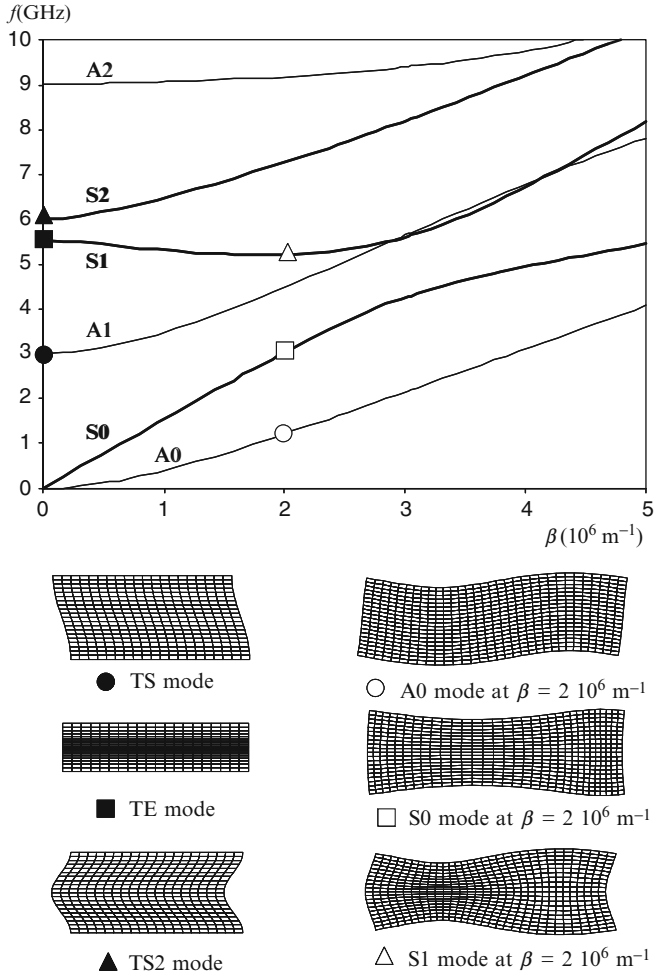

Fig. 4.1 Dispersion curves and displacement fields of Lamb waves propagating in 1-μm-thick AlN plate with top and bottom surfaces grounded and associated displacement fields

associated to some well-known acoustic phenomena[3]: the half-wavelength thickness resonance of the longitudinal wave (or TE resonance mode) corresponding to an ideal BAW resonator for mode S1, the half-wavelength thickness resonance of the horizontally polarized transverse wave (or TS resonance mode) for mode A1, the one wavelength thickness resonance of the horizontally polarized transverse wave (or TS2 resonance mode) for mode S2, etc. Close to the branch point, S1 modes

[3]The location of the resonance modes at $\beta = 0$ is given here for AlN. Their connection to S1 and S2 modes could be different in other materials.

can produce the spurious modes observed in the BAW resonators when resonances of Lamb waves occur due to phase matching of successive reflections on the lateral boundaries of the plate.

The plate can be considered as a first approximation to describe so-called film bulk acoustic resonators or FBAR. To analyse BAW resonators fabricated on a Bragg mirror deposited on a substrate (also called solidly mounted resonators or SMR), the propagation of elastic waves in a plate or a multilayered plate bonded over an isotropic half-space must be considered. This case differs from the plate waveguide which involves total reflection of the partial waves on the boundaries. The infinite half-space is open on the lower side (as in the case of surface waves) making possible the wave radiation in the substrate to infinity, which is equivalent to an energy leakage. For an isotropic plate on an isotropic half-space, a family of non-leaky horizontally polarized shear waves called *Love waves* (which reduce to SH wave when density or stiffness of the substrate goes to zero) can be found when $V_T^{plate} \leq \omega/\beta \leq V_T^{substrate}$. The same modification is found for Lamb waves which become *generalized Lamb waves* in the presence of a substrate. Generalized Lamb waves can only exist when $\omega/\beta \leq V_T^{substrate}$. Their behaviour is rather complicated and depends upon the values of βh (h being the plate thickness) and the ratio $V_T^{plate}/V_T^{substrate}$.

Figure 4.2 displays the dispersion curves of generalized Lamb waves in a 1-µm-thick AlN and zinc oxide (ZnO) plates bonded on a silicon (Si) substrate with some corresponding displacement fields. Different behaviours are observed for the two cases as $V_T^{ZnO} \leq V_T^{Si} \leq V_T^{AlN}$. Only one general Lamb wave is found for the bonded AlN plate, whereas an infinite number of branch solutions is obtained for ZnO (only first two branches are seen in Fig. 4.2). Finally, when βh becomes large, interface waves, also called *Stoneley waves*, localized at the plate-substrate boundary can exist under some conditions. The analysis becomes even more complicated for multilayered plates bonded on a substrate [4].

When designing a BAW resonator, all these elastic waves and dispersion curves have to be considered to explain two- or three-dimensional effects such as spurious modes or energy radiation into the substrate [5–9]. However, these elastic waves can also be utilized to design new acoustic IF and RF MEMS, as discussed in Sects. 4.2.1 and 4.2.2.

4.1.1.3 Elastic Waves in Periodic Media

An elastic periodic medium is constituted by an elementary elastic volume which is periodically repeated along one or several directions. Considering one-dimensional periodicity along x_1, (4.1) becomes

$$\rho(x_1)\frac{\partial^2 u_i}{\partial t^2} = c_{ijkl}(x_1)\frac{\partial^2 u_l}{\partial x_j \partial x_k}, \tag{4.3}$$

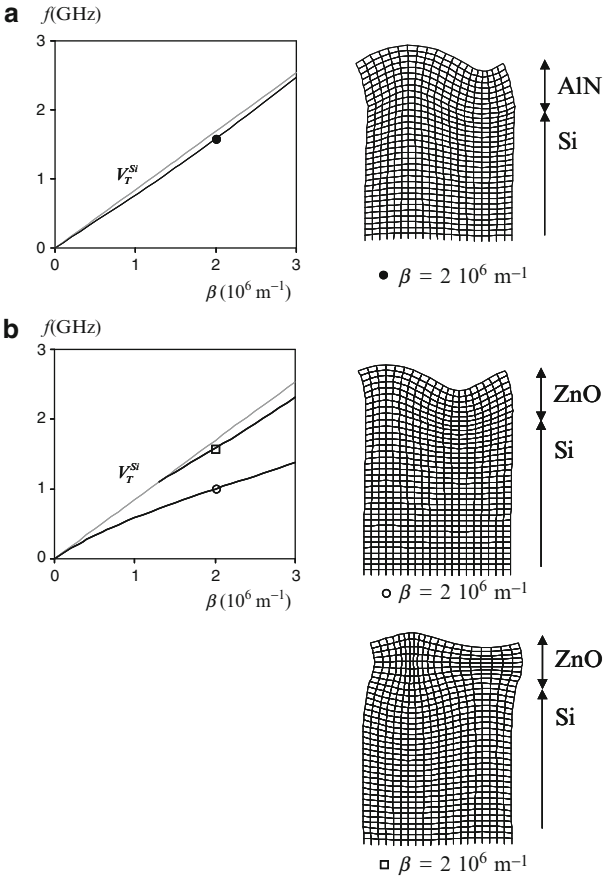

Fig. 4.2 Dispersion curves and displacement fields of generalized Lamb waves propagating in 1-μm-thick AlN (**a**) and ZnO (**b**) plates bonded on semi-infinite silicon substrate with top and bottom plate surfaces grounded

with $\rho(x_1+nd)=\rho(x_1)$ and $c_{ijkl}(x_1+nd)=c_{ijkl}(x_1)$, d being the period of the medium and n an integer. At circular frequency ω, the general solution for wave displacement given by (4.2) implies that

$$u_i(x_1+d,x_2,x_3,t)=u_i(x_1,x_2,x_3,t)\,e^{-j\beta d}, \tag{4.4}$$

where βd appears as an angle which is only known as modulus 2π. Consequently, ω must be a periodic function of β of period $2\pi/d$. The study of the dispersion curve is therefore limited to the interval $[-\pi/d,+\pi/d]$ called first Brillouin zone.

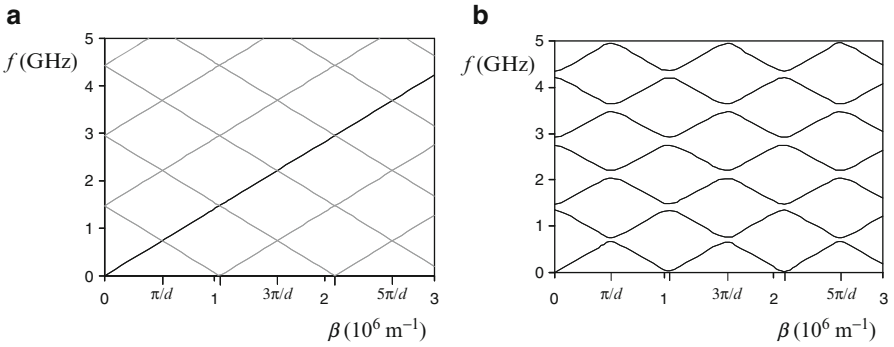

Fig. 4.3 Dispersion curves on longitudinal waves propagating in an infinite isotropic silicon substrate artificially periodized along one direction (**a**) and in infinite stack of alternate Si and SiO_2 layers (**b**). $d = 6\ \mu$m, $h^{Si} = 11d/12$, $h^{SiO2} = d/12$

In an artificially periodized medium,[4] dispersion curves lead to numerous branch crossings at $\beta = n\pi/d$ where n is a positive or negative integer (Fig. 4.3a). Two wave solutions exist for each branch intersection, corresponding to branches centred at different β values. The branches do cross on the dispersion curve as these waves do not interact. If the medium is really periodic, the two waves interact at $\beta = n\pi/d$ and combine into normal propagation modes. Branch splitting and band gap opening are observed (Fig. 4.3b). The phenomena of propagation modes interaction and band gap opening is not limited to Brillouin zone boundaries in periodic media. They happen in any dispersion curve as soon as two branches, corresponding to propagation modes of the same class of symmetry, get close. Thus, in Fig. 4.1, symmetrical (resp. anti-symmetrical) Lamb modes interact without crossing, while mode S and A can intersect without interaction.

The possibility to tailor the dispersion curve by modifying geometry and constitutive materials of the periodical medium has opened a wide field of research on acoustic material engineering. The Bragg mirror, constituted by a stack of alternate high and low acoustic impedance layers and used to isolate the BAW resonator from the substrate in the SMR technology, is a well-known example of one-dimensional periodic structure. Bragg mirror can also be designed to modify the dispersion curves of the guided waves close to the thickness longitudinal resonance in order to improve spurious modes control [5], or to optimize isolation of generalized Lamb waves from substrate and design guided wave IF or RF resonators in an SMR technology [10].

Phononic crystals (PC) are periodic structures designed to exhibit *absolute band gaps*, i.e. frequency bands in which the propagation of elastic waves is forbidden

[4]This case corresponds to a homogeneous medium treated as a periodic medium of period d. The dispersion curve is obtained by superposing homogeneous medium dispersion curves and their replication shifted by any integer (positive or negative) multiple of $2\pi/d$.

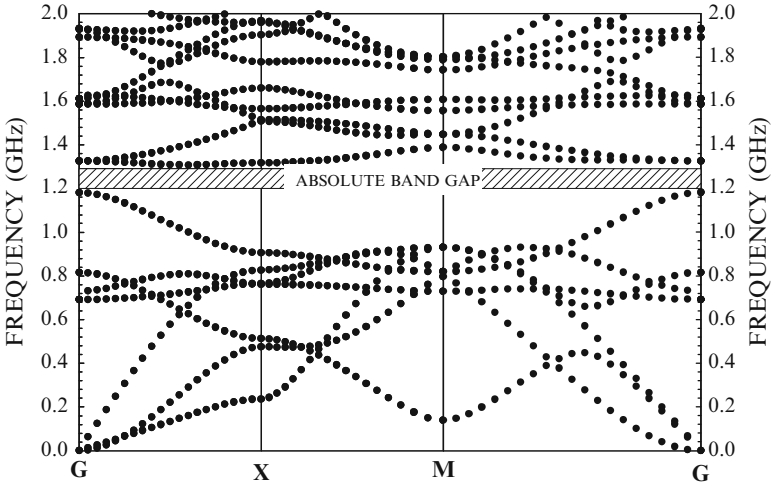

Fig. 4.4 Dispersion curve of elastic waves propagating in a phononic crystal constituted by a square array of cylindrical cavities in a PZT plate

in all directions [11–13]. Works on phononic crystals cover many different fields and applications. In this chapter, focus is on PC properties which could lead to the emergence of new RF MEMS components. Typical dispersion curve of guided elastic waves in a two-dimensional PC plate [14] is displayed in Fig. 4.4. The absolute band gap extends from 1.18 to 1.31 GHz. This property could be used to design a bandstop filter, as described in Sect. 4.3.1, or to isolate very efficiently a MEMS resonator from substrate. However, a most ambitious application is the use the band gap as a basis to select and process the signal propagated by the guided elastic wave through the PC. By adding point defects, linear defects or local resonators in the PC, it becomes possible to realize guiding, multiplexing or demultiplexing of elastic waves, as described in Sect. 4.3.2.

4.1.2 Transduction Mechanisms for RF Devices

This section describes electromechanical energy conversion using materials intrinsic properties which is the most widely used transduction mechanism in acoustic RF MEMS. In insulating materials, mechanical and electrical quantities can be coupled through two different physical effects: piezoelectricity and electrostriction. A detailed presentation of these effects can be found in the following textbooks: [1, 15, 16] for piezoelectricity, [17] for piezoelectricity and electrostriction.

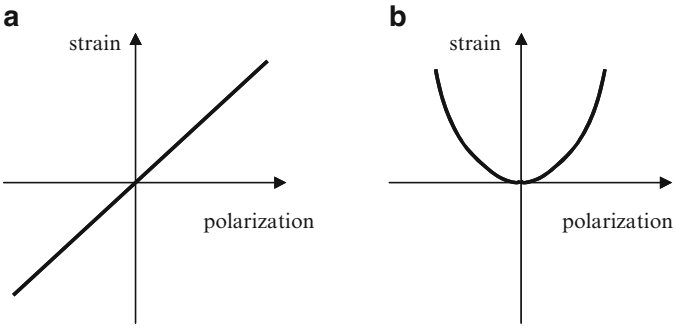

Fig. 4.5 (**a**) Piezoelectric effect, (**b**) electrostrictive effect

4.1.2.1 Piezoelectricity

In some material, strain can induce electrical polarization. This effect is called *direct piezoelectric effect* if it is linear, i.e. if the polarization is proportional to strain and if the sign of polarization is reversed when the sign of strain is reversed (Fig. 4.5a). The direct piezoelectric effect is always accompanied by an *inverse piezoelectric effect*: an external electric field induces a strain of the piezoelectric material. Piezoelectricity is intimately related to microscopic material symmetry. Among the 32 classes of symmetry, piezoelectric crystals are found in 20 non-centrosymmetrical classes. Aluminium nitride (AlN) and zinc oxide (ZnO) are the most commonly used materials in acoustic RF MEMS. They belong to the hexagonal 6-mm class of symmetry which is characterized by a preferred axis (polarization axis conventionally denoted Ox_3) and an isotropic behaviour in the perpendicular plane (O x_1 x_2).

Constitutive equations of piezoelectricity, which are deduced from thermodynamic potentials, are written as

$$\begin{cases} S_{ij} = s^E_{ijkl} T_{kl} + d_{mij} E_m \\ D_m = d_{mij} T_{ij} + \varepsilon^T_{mn} E_n \end{cases}, \tag{4.5}$$

where T_{kl} and E_n are respectively the components of the stress tensor and of the electric field vector which are chosen as independent variables. S_{ij} and D_m are respectively the components of the strain tensor and of the electric displacement vector. s^E_{ijkl}, d_{mij} and ε^T_{mn} denote the components of the elastic compliance tensor at constant electric field, the piezoelectric tensor and the dielectric permittivity tensor at constant stress, respectively. Subscripts i to n vary from 1 to 3. Equation (4.5) emphasizes the close relationship, in a piezoelectric RF device, between the applied electric field, the induced strain and the orientation of the material. In a classical AlN or ZnO TE BAW resonator, polarization axis x_3 is orientated along thickness. Electrodes on top and bottom surfaces of the piezoelectric film generate an electric

field along x_3 which mainly induces a strain (and therefore a vibration) along x_3 due to the symmetry of the piezoelectric tensor. At the design stage, the choice of polarization direction, electrode geometry and resonator geometry must be optimized to maximize excitation of a specific elastic wave and the consequent resonance mode, e.g. the half-wavelength longitudinal thickness mode in a TE resonator.

The electromechanical coupling factor is used to evaluate the efficiency of this overall conversion. For a transducer, the *electromechanical coupling factor* k is defined as the ratio of the mutual energy W_{mut} (work transformed from one form of energy to another) to the geometric mean of the work done on the blocked electrical impedance W_e and the work done on the open-circuit mechanical impedance W_m

$$k = \frac{W_{mut}}{\sqrt{W_e W_m}}. \tag{4.6}$$

If the work is done only on one port, the electrical port in the case of an acoustic RF MEMS, k is defined as the square root of the ratio of the work done on the electrical port transduced into mechanical energy to the total work done on the electrical port

$$k = \sqrt{\frac{W_{mut}}{W_e}}. \tag{4.7}$$

Considering the Butterworth–van Dyke representation of a transducer on which a quasistatic voltage V is applied (Fig. 4.6), the works are $W_e = (C_0 + C_M)V^2/2$ and $W_{mut} = C_M V^2/2$, leading to

$$k = \sqrt{\frac{C_M}{C_0 + C_M}}. \tag{4.8}$$

Defining the transducer series resonance frequency f_s and parallel resonance frequency f_p as

$$f_s = \frac{1}{2\pi\sqrt{L_M C_M}}, \tag{4.9a}$$

$$f_p = \frac{1}{2\pi\sqrt{L_M (C_0 + C_M)}}, \tag{4.9b}$$

k can then also be written as

$$k = \sqrt{1 - \frac{f_s^2}{f_p^2}}. \tag{4.10}$$

The electromechanical coupling coefficient can be evaluated by measuring series and parallel resonances on the electrical port. This evaluation is exact at low frequency or for some particular transducer geometries and modes (length resonance of bars, thickness resonance of plates, radial resonance of thin rings or thin spheres, etc.).

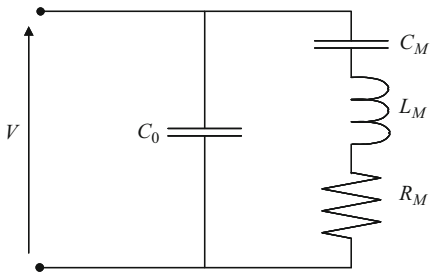

Fig. 4.6 Lumped parameter Butterworth van Dyke circuit representation of a transducer. C_0, C_M, L_M and R_M are respectively the blocked capacitance, the motional capacitance, the motional inductance and the motional resistance

Table 4.1 Coupling coefficient of thin piezoelectric film [18]

Material	AlN	ZnO	PT	PZT	KNO	KLN	LNO[a]	LTO[b]
$k_t^2(\%)$	6.5	9.5	19	23	45	25	2.8	3.7

[a]Lithium niobate, $LiNbO_3$
[b]Lithium tantalate, $LiTaO_3$

In BAW resonators, the coupling coefficient is an important parameter as it determines approximately the bandwidth of the bandpass filter build with these resonators. Table 4.1 displays the coupling coefficient k_t^2 (corresponding to TE mode) for various piezoelectric materials grown in thin films. In AlN or ZnO BAW resonators, k_t^2, from 6.5% to 9.5%, is well suited for emission and reception bands of mobile phone communication standards. Much larger bandwidths could be obtained with potassium lithium niobate (KLN), lead titanate (PT), lead zirconate titanate (PZT) or potassium niobate (KNO), if they become technologically compatible without degradation of their properties.

Other modes than the thickness longitudinal mode could be exploited. Progress in this field requires new geometries of resonator and/or electrodes or new technological processes for thin piezoelectric layer fabrication. Thus, deposition of tilted AlN or ZnO layers has drawn lot of interest these last years in order to develop TS resonators. Perspectives on TS BAW resonators are given in Sect. 4.2.1.

4.1.2.2 Electrostriction

When an insulating material is submitted to an electric field, *electrostriction* produces a strain which is a quadratic function of polarization (Fig. 4.5b). At the microscopic level, this effect, similar to the inverse piezoelectric effect, does not require specific crystal symmetry. Electrostriction is a second order effect, which exist in all materials but is often negligible. Barium titanate (BTO) and lead magnesium niobate (PMN) are bulk materials exhibiting a large electrostrictive effect due to their high dielectric permittivity. Strontium titanate (STO) and barium strontium titanate (BST) are the most studied electrostrictive material in acoustic RF MEMS.

Constitutive equations of electrostriction are written as

$$\begin{cases} S_{ij} = s^D_{ijkl} T_{kl} + Q_{ijmn} D_m D_n \\ E_m = -2 Q_{ijmn} D_n T_{ij} + \beta^T_{mn} D_n \end{cases}, \tag{4.11}$$

where the stress tensor and the electric displacement are chosen as independent variables. s^D_{ijkl}, Q_{ijmn} and β^T_{mn} denote the components of the elastic compliance tensor at constant electric displacement, the electrostrictive tensor and the dielectric impermeability tensor at constant stress respectively.

As electrostriction is a quadratic effect, the electrostrictive BAW resonator is submitted to a DC electric field with superposed small-signal AC field. Equivalent linearized elastic, piezoelectric and dielectric constants, depending upon DC field, can be derived from (4.11) [19]. As a consequence, the DC electric field can control the coupling coefficient, opening the way for the design of switchable and tunable resonators described in Sect. 4.2.3.

4.2 Acoustic RF Resonators and Bandpass Filters

Resonators are the most studied acoustic RF devices. As elastic wavelengths are much smaller than electromagnetic wavelengths at a given frequency, resonators can be miniaturized by using elastic waves. In the RF range, the elastic wavelength in usual materials (0.1–1 μm at 1 GHz) coincides with the dimensions of the objects that can be fabricated with the MEMS technology. As the classical thickness extensional BAW resonator is already discussed in details in other chapters of this book, this section focuses on emerging RF devices using shear and guided elastic waves and on tunable TE BAW resonators.

4.2.1 Thickness-Shear Resonators

Thickness-shear (TS) resonators, which exploit the half-wavelength thickness resonance of the transverse or shear wave (Fig. 4.1), are mainly studied for sensing applications in liquids. However, there is some interest to consider their use in RF filters. As the transverse or shear wave is slower than the longitudinal wave and piezoelectric layers with "good properties" deposited by sputtering are thickness limited, TS resonators could address a lower frequency band (such as GSM 900) than the TE resonator. In addition, TS and TE resonators could be co-integrated on a single piezoelectric material layer to address different communication standards. Finally, the coupling coefficient can be theoretically larger for TS resonator than for TE resonator in some piezoelectric materials such as ZnO.

In AlN and ZnO, the TS thickness can be excited electrically with plain top and bottom electrodes when the polarization axis of the piezoelectric material is

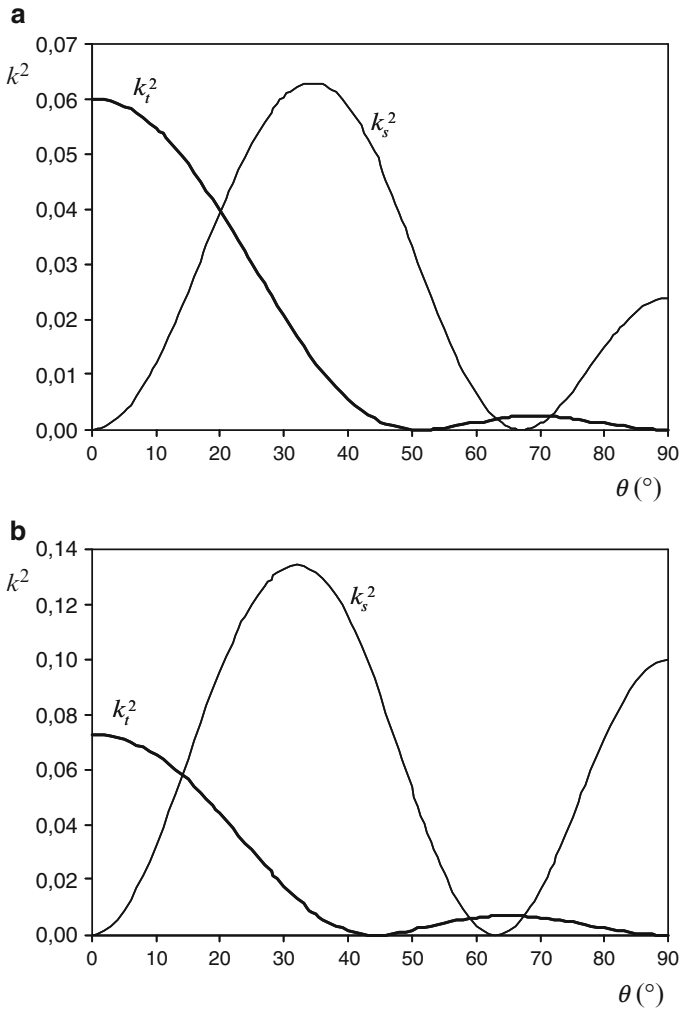

Fig. 4.7 Coupling coefficients versus polarization axis orientation for TE (k_t^2) and TS (k_s^2) in AlN and ZnO thin films. θ is the angle between the polarization axis and the film thickness direction. Bulk material characteristics are used and electrodes are neglected

adequately tilted with respect to the direction normal to the layer. The theoretical values (calculated from bulk material properties) of k_t^2 for TE mode and k_s^2 for TS mode versus tilt angle θ are displayed in Fig. 4.7. Maximum values for ZnO (resp. AlN) TS resonator are $k_s^2 = 13.4\%$ when $\theta = 32^\circ$ (resp. $k_s^2 = 6.3\%$ when $\theta = 34^\circ$). TE modes are also coupled at these angles. The longitudinal coupling coefficient goes to zero at $\theta = 44^\circ$ for ZnO (resp. $\theta = 51^\circ$ for AlN) where $k_s^2 = 9.3\%$ (resp. $k_s^2 = 3\%$). Characteristics of recently fabricated TS resonators with titled piezoelectric material are given in Table 4.2. Different methods of fabrication are

Table 4.2 Measured characteristics of recent TS resonators

	Link et al. [21]	Martin et al. [20]	Bjurström et al. [22]	Yanagitani et al. [23]	Milyutin and Muralt [24]
Material	ZnO	AlN	AlN	ZnO	AlN
Layer and orientation	One layer tilted 18°	One layer tilted 6° (001)	One layer tilted 28°–32° (103)	Two layers tilted 90° ($11\bar{2}0$) of opposite polarization	One layer not tilted (001)
Electrodes	Plain	Plain	Plain	Plain	Interdigitated
Resonator type	SMR	SMR	FBAR	FBAR	SMR
f_s (GHz)	0.85	5	1.64	0.25	1.85
k_s^2 (%)	1.6	0.5	2	3.3	–
Q	312	110	350	–	1,000

used to obtain the tilted orientation: substrate holder tilted during sputtering [20], blind constituting an additional anode near the substrate [21] and use of anisotropic deposition flux far from the target centre [22]. Measured coupling coefficients are much smaller than theoretical values. Quality factors,[5] around 300, remain too low for RF filter applications.

Other solutions have been developed to couple TS mode (Table 4.2): in reference [23] a larger shear coupling coefficient is obtained with plain electrodes by depositing two layers of ZnO with opposite polarization axes aligned along substrate plane; in reference [24] classical thickness oriented AlN with interdigitated electrodes is used to reach a quality factor of 1,000 and $Q \cdot f$ product of 1.8×10^{12}.

4.2.2 Guided Acoustic Wave Resonators and Filters

In guided acoustic wave resonators, the resonance effect is obtained by in-phase looping of Lamb waves or generalized Lamb waves. The guided wave propagation is parallel to the plate surface, and the resonance occurs whether along one specific direction ("lateral" resonator) or in the plate plane ("contour" resonator). The resonance frequency of guided acoustic wave resonators is determined by the resonator lateral dimensions and the guided wavelength. Typical geometries of guided wave resonators are displayed in Fig. 4.8. They mainly differ by their lateral boundaries (plate/air interface or Bragg mirror). In the process, resonator critical dimension is obtained by etching and not by film deposition as in BAW resonator. A wider range of dimensions is therefore available, making possible the realization of guided wave resonators in IF and RF band with the same technology, as well as and their co-integration with BAW resonators [25].

The characteristics of recently fabricated guided wave resonators exploiting S0 Lamb wave are given in Table 4.3. All are of FBAR type. Resonance frequencies range from 15 MHz to 1.25 GHz. Effective coupling coefficients squared (k^2) are low, between 0.4% and 0.8%, as expected for the S0 wave. Q-factors range from 1,400 to 6,100, and high $Q \cdot f$ products around 1.8×10^{12} can be reached [26]. Guided wave solidly mounted resonators have not been fabricated yet, but Bragg reflectors have been designed to isolate guided waves and resonator characteristics have been theoretically evaluated [10].

Guided wave resonators have already been implemented in several communication devices:

- Compact narrowband IF and RF channel-select filters using acoustically coupled guided wave resonators [30, 31]. Such devices could replace off-chip SAW filters and LC components and enable total integration of an heterodyne RF receiver

[5] Quality factors given in this chapter are in-air measured values at series frequency f_s unless specified otherwise.

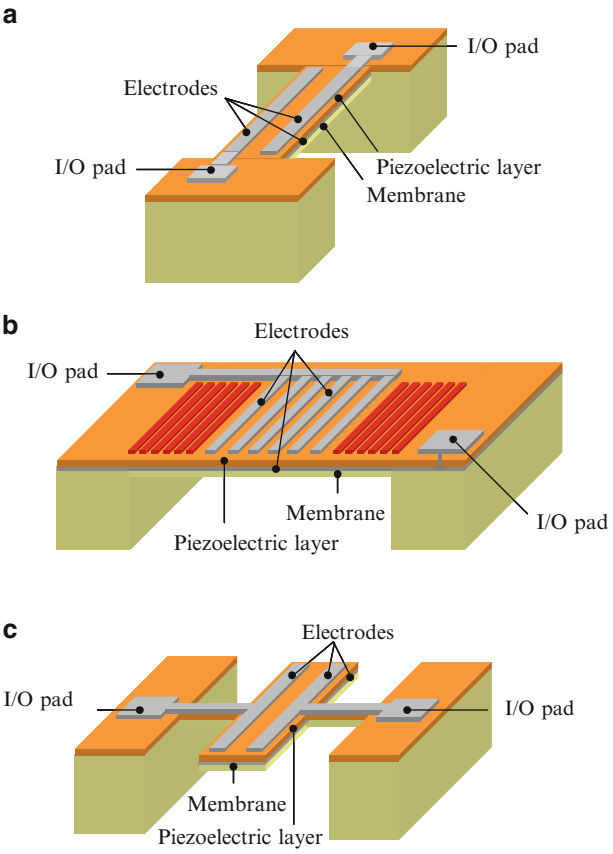

Fig. 4.8 Typical geometries of guided wave resonators: (**a**) lateral resonator with etched boundaries, (**b**) lateral resonator with Bragg mirrors, (**c**) contour resonator suspended by anchors

architecture using simultaneously IF and RF filters. They are also well suited for low power, low to medium bandwidth applications such as IEEE 802.15.4 (Zigbee).

- Low phase-noise reference UHF oscillators in which guided wave resonators replace electromechanical resonators as they exhibit lower motional impedance and operate without DC polarization [32].

4.2.3 Tunable Thickness Extensional Resonators

Mobile terminals become more and more multi-standard and multimode. Conventional approach, based on switching between different circuits dedicated to a specific standard, is not effective in terms of cost and miniaturization. The trend is to use

Table 4.3 Measured characteristics of recent S0 guided acoustic wave resonators

	Volatier et al. [25]	Yantchev and Katardjiev [27]	Stephanou and Pisano [26]	Piazza et al. [28]	Ho et al. [29]
Material	AlN	AlN	AlN	AlN	ZnO
Resonator type	FBAR[a]	FBAR	FBAR	FBAR	FBAR
Lamb wave type	S0	S0	S0	S0	S0
Mode type	Lateral	Lateral	Lateral	Contour	Contour
Mode order[b]	1–7	40	9	(1,0) and (0,1)	(1,0) and (0,1)
Lateral boundary conditions	Suspended along shorter sides	Strip reflectors	Suspended by anchors	Suspended by anchors	Suspended by anchors
f_s (GHz)	0.02–0.28	0.89	1.28	0.02 and 0.085	0.015 and 0.09
k^2 (%)	up to 0.8	0.56	0.41	0.52 and 0.68	–
Q	up to 2,000	3,100	1,400	2,000–3,000	6,100 and 3,400

[a] Co-integrated with TE FBAR at 2.5 GHz

[b] Mode order is denoted by l for lateral resonator where l is the number of half-wavelengths along the direction of guided wave propagation and by (n,m) for contour resonator where n (resp. m) is the number of half-wavelengths along resonator length (resp. width)

reconfigurable components and thus to develop filters with tunable central frequency and bandwidth. For example, 9% (resp. 16%) relative frequency shift is required to switch from DCS to PCS (resp. to W-CDMA). Tunable resonators could also be used to reduce dispersion of BAW resonance frequencies (due to variations of deposited layer thickness) after fabrication and thus to avoid the complicated laser ablation process. In that case, $\pm 1.1\%$ relative frequency shift is needed.

Although tunable resonators have been realized by associating conventional BAW resonators with variable capacitances, this section focuses on tunability obtained by exploiting intrinsic properties of ferroelectric and paraelectric materials. Ferroelectric materials are characterized by a spontaneous electric polarization and an hysteresis loop (due to domain switching) in the polarization curve. Paraelectric materials require an external electric field for polarization and do not exhibit hysteresis. They can be described by the theory of electrostriction. In both cases, the application of a DC bias produces a change of small-signal effective elastic, "piezoelectric" and dielectric properties, and therefore a change of resonator characteristic frequencies and coupling coefficient.

Table 4.4 summarizes measured characteristics of recent (ferroelectric) PZT, (paraelectric) STO and BST made TE resonators. Frequency shifts obtained under DC bias are large enough to correct dispersion caused by layer thickness variations but not to switch between RF standards. In most cases, shifts vary differently for

Table 4.4 Measured characteristics of recent tunable TE resonators

	Schreiter et al. [33]	Schreiter et al. [33]	Zinck et al. [34]	Conde and Muralt [35]	Berge et al. [36]	Volatier et al. [19]
Material	PZT 25/75	PZT 58/42	PZT 52/48	PZT 53/47	BST 25/75	STO
Resonator type	SMR	SMR	FBAR	FBAR	SMR	SMR
f_s at low bias (GHz)	2.06	1.87	1.55	0.78	4.4	2.2
Max. bias field (MV/m)	+28/−28	+32/−32	+25/−25	+20/−20	+47	+70
Max. relative f_s shift (%)	+0.3/−0.1	+1.5/−1.9	−1.4/+0.7	−3.4/+1.8	−1.6	−1.6
Max. relative f_p shift (%)	+0.5/−1.8	+0.1/−0.5	−0.1/+0.7	−2.1/+1.1	−1.3	−0.6
Hysteresis	Yes	Yes	Yes	Yes	No	No
Max. k_t^2 (%)	4.4	9.0	7.3	9.0	0.7	3.0
Bias field at zero k_t^2 (MV/m)	+16.8/−6.5	+2.6/−2.6	+2.5/−5.0	+1.2/−7.0	0.0	0.0
Q	220	220	85 (at f_p)	50	120	50

series and parallel resonances, as a consequence of coupling coefficient variation with DC bias. To use such resonators, new filter architectures would be required to shift central frequency without bandwidth modification.

Conde and Muralt [35] have found, for 53/47 PZT resonator, a range of DC bias where k_t^2 was kept constant around 7%. A bandpass filter using these resonators would theoretically have a 1.3% relative shift of the central frequency without modification of the bandwidth. Quality factors of tunable TE resonators remain low. In PZT, the low Q value is attributed to the ferroelastic domain wall relaxation which occurs close to BAW operating frequency and can therefore not be avoided [18]. Paraelectric materials do not have this limitation. BST and STO resonators' quality factors and coupling coefficients are currently limited by the quality of the active layer. Much higher values are expected from improved processes.

4.3 Acoustic RF Devices Based on Phononic Crystals

Phononic crystals (PC) have emerged recently as candidates for RF signal processing applications, as photonic crystals did previously in optics [37]. Properties of guided elastic wave dispersion curves in PC (described in Sect. 4.1.1.3) can be exploited with propagating or stationary waves. In the first case, the guided wave propagates through the PC, and emitting/receiving IDT are used as electrical input/output. In such devices, dispersion curve properties are easily transposed into a specific RF function, but high insertion loss is expected. Devices involving stationary waves may have low insertion loss when a resonance is used, but the design of a specific function from dispersion curves is not straightforward.

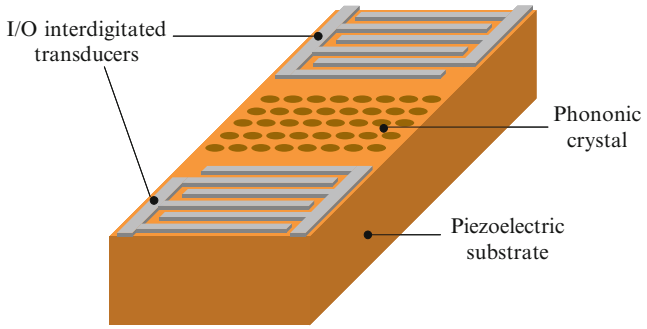

Fig. 4.9 Typical geometry of a phononic crystal bandstop filter using surface acoustic waves

Table 4.5 Measured characteristics of PC-based SAW bandstop filters

	Wu et al. [39]	Benchabane et al. [40]	Benchabane et al. [40]	Assouar et al. [41]
Material	Silicon[a]	y-cut LNO	y-cut LNO	z-cut LNO
Lattice geometry	Square	Square	Square	Square
Hole geometry	Cylinder	Cylinder	Cylinder	Hexagon
Hole volumic fraction (%)	38.5	64	64	28
Propagation direction	ΓX	ΓX	ΓM	ΓY
Central frequency (MHz)	200	220	215	297
Bandwidth (%)	5	60	23	15
Insertion loss (dB)	65	20	25	52
Rejection level (dB)	10	10	20	18

[a] To generate and detect elastic waves, thin ZnO films are deposited on silicon substrate on each side of the silicon PC

The very few PC-based devices fabricated until now involve surface acoustic waves. Some theoretical results as well as measurements on acoustic waves in fluid PC are also reported in this section.

4.3.1 Resonators and Filters

An elastic wave propagating through a PC is evanescent in the band gap. A *bandstop filter* function is therefore realized between input and output interdigitated electrodes (IDT). Several PC IF bandstop filters have been realized with surface acoustic waves propagating through a lattice of holes in silicon or LNO substrates (Fig. 4.9 and Table 4.5). Relative bandwidths (BW) range from 11% to 27%. Rejection level (RL), comprised between 10 and 20 dB, is limited by transmission of leaky SAW waves. As expected, insertion loss (IL) is high (between 20 and 52 dB). Simulations made for generalized Lamb waves propagating in a square lattice of cylindrical holes in a PZT layer bonded on silicon substrate give similar results: BW = 24%, RL = 18 dB, IL = 25 dB for a hole volumic fraction of 70% [38].

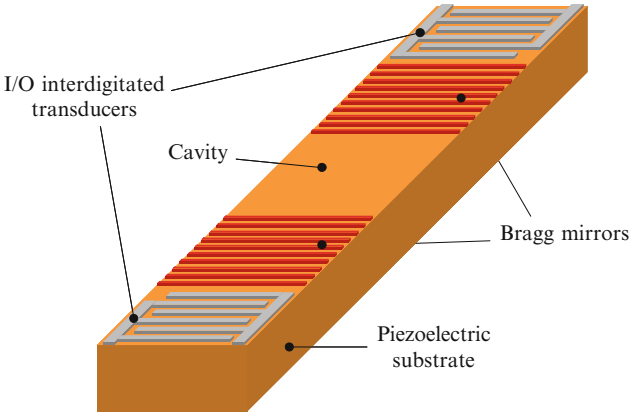

Fig. 4.10 Typical geometry of Fabry–Perot resonator using surface acoustic waves

Acoustic *bandpass filters* have been realized by adding defects in a PC having a very large band gap [42]. In a lattice of PVC cylinders surrounded by air, a waveguide is created by removing a row of cylinders along the propagation direction. Such waveguide transmits very efficiently the waves falling inside the stopband. Resonant filtering is obtained by coupling the waveguide to a side branch resonator. *Notches* appear in the transmission spectrum of the waveguide at the resonance frequencies of the branch resonator. Waveguide BW varies between 25% and 60% with high IL. For notches, BW is typically 2% and RL 5–10 dB. This principle could be extended to elastic waves.

Two PC can be combined via an acoustic cavity to realize a *Fabry–Perot resonator* [43]. A one-dimensional acoustic Fabry–Perot resonator is constituted by two Bragg mirrors delimiting an acoustic cavity (Fig. 4.10). Transducers are located outside this structure for acoustic wave emission and reception. The transmission of acoustic wave through the structure is only possible in very narrow frequency ranges corresponding to resonant propagation through the Bragg mirrors. Due to the interference between the Bragg mirrors, acoustic wave amplitude in the cavity is amplified and low IL can be obtained. Using surface acoustic waves, a resonator at 98.7 MHz with $Q \approx 25{,}000$ has been realized. The same principle could be applied with bulk or guided acoustics waves.

4.3.2 Multiplexers and Demultiplexers

Acoustic multiplexers and demultiplexers are based on resonant tunnelling phenomenon. A device is composed of two parallel waveguides and a resonator between the waveguides acting as coupling element. Figure 4.11 displays an extrapolation of this concept using surface acoustic waves. At some frequencies, an energy transfer

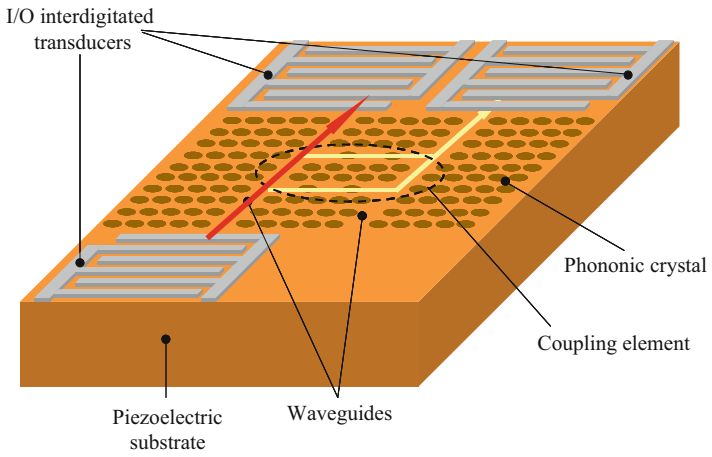

Fig. 4.11 Schematic representation of a PC demultiplexer using surface acoustic waves

occurs between the two waveguides through the resonance modes of the coupling element. This effect, already known for photonic crystals, has been experimentally demonstrated for acoustic waves in a square lattice of composed steel cylinders [44]. BW of the transferred signal is around 1% with amplitude of –10 dB in the input waveguide and −3 dB in the output waveguide.

4.4 Conclusion

Most of recent advances on new acoustic RF MEMS devices in research laboratories are based on other types of acoustic waves than bulk or surface acoustic waves and/or other transduction phenomena than piezoelectricity. Guided acoustic wave resonators and filters exhibit some interesting characteristics and could be rapidly utilized in IF and RF circuits. Much research has been conducted on thickness-shear and tunable resonators, but physical properties of active materials used in these components have still to be improved. Perspectives have been opened up by phononic crystal-based components which are at an early stage of development.

References

1. B.A. Auld, *Acoustic Fields and Waves in Solids* (Krieger Publishing Company, Malabar, Florida, 1990)
2. D. Royer, E. Dieulesaint, *Elastic Waves in Solids* (Springer, Berlin, 2000)
3. L. Brillouin, *Wave Propagation in Periodic Structures* (McGraw-Hill, New York, 1946)
4. V. Zhang, B. Dubus, B. Collet, M. Destrade, J. Acoust. Soc. Am. **123**, 1972 (2008)

5. G.G. Fattinger, S. Marksteiner, J. Kaitila, R. Aigner, in *Proceedings IEEE Ultrasonics Symposium*, 2005, pp. 1175–1178
6. R. Ruby, J. Larson, C. Feng, S. Fazzio, in *Proceedings IEEE Ultrasonics Symposium*, 2005, pp. 217–220
7. D. Ekeom, B. Dubus, A. Volatier, in *Proceedings IEEE Ultrasonics Symposium*, 2006, pp. 1474–1477
8. J. Kaitila, in *Proceedings IEEE Ultrasonics Symposium*, 2007, pp. 120–129
9. R. Ruby, in *Proceedings IEEE Ultrasonics Symposium*, 2007, pp. 1029–1040
10. I. Koné, B. Dubus, L. Buchaillot, A. Reinhardt, F. Casset, M. Aïd, J.F. Carpentier, P. Ancey, in *Proceedings IEEE Frequency Control Symposium*, 2008, pp. 581–585
11. J.O. Vasseur, B. Djafari-Rouani, L. Dobrzynski, P.A. Deymier, J. Phys. Condens. Matter **9:7327** (1997)
12. D. Garcia-Pablos, M. Sigalas, F.R.M. d. Espinoza, M. Torres, M. Kafesaki, N. Garcia, Phys. Rev. Lett. **84:4349** (2000)
13. S.X. Yang, J.H. Page, Z.Y. Liu, M.L. Cowan, C.T. Chan, P. Sheng, Phys. Rev. Lett. **93**, 024301 (2004)
14. J.O. Vasseur, A.C. Hladky-Hennion, B. Djafari-Rouhani, F. Duval, B. Dubus, Y. Pennec, J. Appl. Phys. **101**, 111904 (2007)
15. D.A. Berlincourt, D.R. Curran, H. Jaffe, in *Physical Acoustics, Principles and Methods*, vol. 1A, ed. by W.P. Mason (Academic, New York, 1964)
16. O.B. Wilson, *Introduction to the Theory and Design of Sonar Transducer* (Peninsula Publishing, Los Altos, 1988)
17. W.P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (D. Van Nostrand Company, Princeton, 1950)
18. P. Muralt, J. Antifakos, M. Cantoni, R. Lanz, F. Martin, in *Proceedings IEEE Ultrasonics Symposium*, 2005, pp. 315–320
19. A. Volatier, E. Defaÿ, M. Aïd, A. N'hari, P. Ancey, B. Dubus, Appl. Phys. Lett. **92**, 032906 (2008)
20. F. Martin, M.E. Jan, S. Rey-Mermet, B. Belgacem, D. Su, M. Cantoni, P. Muralt, IEEE Trans. Ultrason. Ferroelect. Freq. Contr. **53**, 1339 (2006)
21. M. Link, M. Schreiter, J. Weber, R. Primig, D. Pitzer, G. Reinhardt, IEEE Trans. Ultrason. Ferroelect. Freq. Contr. **53**, 492 (2006)
22. J. Bjurström, G. Wingqvist, I. Katardjiev, IEEE Trans. Ultrason. Ferroelect. Freq. Contr. **53**, 2095 (2006)
23. T. Yanagitani, M. Kiuchi, M. Matsukawa, Y. Watanabe, IEEE Trans. Ultrason. Ferroelect. Freq. Contr. **54**, 1680 (2007)
24. E. Milyutin, P. Muralt, in *Proceedings IEEE Frequency Control Symposium*, 2008
25. A. Volatier, G. Carruyer, D.P. Tanon, P. Ancey, E. Defaÿ, B. Dubus, in *Proceedings IEEE Ultrasonics Symposium*, 2005, pp. 902–905
26. P.J. Stephanou, A.P. Pisano, in *Proceedings IEEE Ultrasonics Symposium*, 2006, pp. 2401–2404
27. V. Yantchev, I. Katardjiev, in *Proceedings IEEE Ultrasonics Symposium*, 2005, pp. 1580–1583
28. G. Piazza, P.J. Stephanou, A.P. Pisano, J. Microelectromech. Syst. **15**, 1406 (2006)
29. G.K. Ho, R. Abdolvand, A. Sivapurapu, S. Humad, F. Ayazi, J. Microelectromech. Syst. **17**, 512 (2008)
30. M. Desvergne, E. Defaÿ, D. Wolozan, M. Aïd, P. Vincent, A. Volatier, Y. Deval, J.B. Bégueret, in *Proceedings European Solid State Device Research Conference (ESSDERC)*, 2007, pp. 358–361
31. R. Abdolvand, F. Ayazi, in *Proceedings IEEE/MTT-S International Microwave Symposium*, 2007, pp. 509–512
32. H.M. Lavasani, R. Abdolvand, F. Ayazi, in *Proceedings IEEE Custom Integrat. Circ. Conf.*, 2007, pp. 599–602
33. M. Schreiter, R. Gabl, D. Pitzer, R. Primig, W. Wersing, J. Euro. Ceram. Soc. **24**, 1589 (2004)

34. C. Zinck, E. Defaÿ, A. Volatier, G. Caruyer, D.P. Tanon, L. Figuire, in *Proceedings IEEE Ultrason. Ferroelectr. Freq. Control. joint 50th Anniv. Conf.*, 2004, pp. 29–32
35. J. Conde, P. Muralt, IEEE Trans. Ultrason. Ferroelect. Freq. Contr. **55**, 1373 (2008)
36. J. Berge, A. Vorobiev, W. Steichen, S. Gevorgian, IEEE Microwave Wireless Comm. **17**, 655 (2007)
37. J.M. Lourtioz, *Photonic Crystals: Towards Nanoscale Photonic Devices* (Springer, New York, 2005)
38. A.C. Hladky-Hennion, J. Vasseur, B. Dubus, F. Duval, C. Granger, Y. Pennec, B. Djafari-Rouhani, B. Morvan, in *Proceedings IEEE Ultrasonics Symposium*, 2007, pp. 620–623
39. T.T. Wu, L.C. Wu, Z.G. Huang, J. Appl. Phys. **97**, 094916 (2005)
40. S. Benchabane, A. Khelif, J.Y. Rauch, L. Robert, V. Laude, Phys. Rev. E **73**, 065601(R) (2006)
41. B.M. Assouar, B. Vincent, H. Moubchir, IEEE Trans. Ultrason. Ferroelect. Freq. Contr. **55**, 273 (2008)
42. J.O. Vasseur, P.A. Deymier, M. Beaugeois, Y. Pennec, B. Djafari-Rouani, D. Prevost, Zeitschrift fur Kristallographie, pp. 829–835 (2005)
43. S.G. Alekseev, Y.V. Gulyaev, G.D. Mansfeld, V.I. Pustovoit, V.F. Dmitriev, in *Proceedings IEEE Ultrasonics Symposium*, 2005, pp. 2124–2127
44. Y. Pennec, B. Djafari-Rouhani, J.O. Vasseur, H. Larabi, A. Khelif, A. Choujaa, S. Benchabane, V. Laude, Appl. Phys. Lett. **87:261912** (2005)

Part II
MEMS-Based Circuits

Chapter 5
The Design of Low-Power High-Q Oscillators

Eric A. Vittoz

Abstract This chapter starts with a general theory of high-Q oscillators that includes all nonlinear effects and a basic approach of phase noise. Then, three important circuit architectures that are compatible with MEMs or quartz crystal resonators are discussed. For each of them, a linear analysis is first carried out to obtain the critical transconductance for oscillation and the amount of frequency pulling from the mechanical resonance of the resonator. The dependency of the amplitude on the bias current is then derived from the transfer characteristics of the active transistor(s) by using the concept of transconductance for the fundamental component. A more detailed noise analysis is followed by the presentation of practical implementations, with one example of amplitude limitation circuitry. The respective merits of the three approaches are compared in a last section.

5.1 Introduction

A quartz crystal or a MEM resonator is basically a passive mechanical device driven electrically by means of an electromechanical transducer. An oscillator is built by combining it with an electronic circuit that provides the necessary energy of oscillation and compensates the losses. The frequency of oscillation should be defined by the precise resonant frequency of the resonator, with a minimum influence of the circuit. This is made easier by the value of the quality factor Q of the resonator that is much larger than that of a standard electrical resonator.

For the usual case of an electrostatic or a piezoelectric transducer, the electrical circuit equivalent to the mechanical oscillator is an R_m, L_m, C_m *series resonator*. The

E.A. Vittoz (✉)
Ecole Polytechnique Féderale de Lausanne (EPFL),
1015 Lausanne, Switzerland
e-mail: evittoz@ieee.org

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_5,

limited amount of electromechanical coupling results in an unavoidable capacitor C_0 of large value, connected across the series resonator. Thus, the device also exhibits a high-impedance parallel resonance (or antiresonance) when the oscillatory current circulates in the R_m, L_m, C_m, C_0 loop.

High-Q and parallel capacitance C_0 are the two main differences with respect to a normal electrical series resonator. In this chapter, three well-known oscillators will be analyzed and modified to take into account the presence of C_0 while taking advantage of the high value of Q. These are the single-transistor Pierce circuit [1], the symmetrical oscillator for parallel resonance, and the symmetrical oscillator for series resonance. A general theory of high-Q oscillators that includes all nonlinear effects will first be explained.

Most of the material presented in this chapter is extracted from a more extended analysis of circuits for quartz and MEMs oscillators [2].

5.2 General Theory of High-Q Oscillators

5.2.1 Splitting of the Oscillator for Nonlinear Analysis

The general form of an oscillator based on an electromechanical resonator with piezoelectric or electrostatic transduction and a sustaining circuit is illustrated in Fig. 5.1a.

The motional impedance Z_m is an $R_m L_m C_m$ series resonant circuit that corresponds to the mechanical part of the resonator. The mechanical resonant frequency ω_m and quality factor Q are given by

$$\omega_m = 1/\sqrt{L_m C_m} \quad \text{and} \quad Q = \frac{1}{\omega_m C_m R_m} = \frac{\omega_m L_m}{R_m}. \tag{5.1}$$

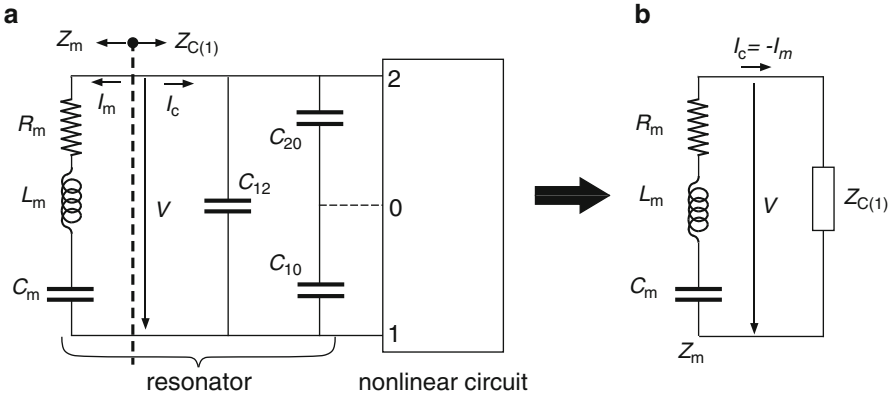

Fig. 5.1 General form of an oscillator based on a electromechanical resonator with piezoelectric or electrostatic transduction: (**a**) combination of the resonator with a nonlinear circuit; (**b**) splitting into motional impedance Z_m and circuit impedance $Z_{c(1)}$

The motional inductance L_m is proportional to the mass of the resonator, whereas the motional capacitance C_m is proportional to the inverse of its stiffness. The motional resistance R_m represents the mechanical losses.

There is always a capacitance C_{12} in parallel with the series resonator because only a (small) part of the electrical energy delivered to the resonator by applying a voltage V across it is transformed into mechanical energy. The ratio C_m/C_{12} represents the electromechanical coupling. For piezoelectric transducing, it is limited to the intrinsic coupling coefficient of the material, which is about 1% for quartz. For electrostatic transducing, it depends on the DC voltage applied to the transducing capacitor.

C_{10} and C_{20} are capacitances with respect to the packaging case, which is usually grounded. If the resonator is used as a dipole (without the connection in dotted line) or if the circuit is perfectly symmetrical, the three electrical capacitances can be lumped into a single parallel capacitor:

$$C_0 \triangleq C_{12} + \frac{C_{10}C_{20}}{C_{10}+C_{20}}. \tag{5.2}$$

The motional current I_m is proportional to the velocity of oscillation. The mechanical energy of oscillation is

$$E_m = \frac{L_m|I_m|^2}{2} = \frac{|I_m|^2}{2\omega_m^2 C_m} = \frac{QR_m|I_m|^2}{2\omega_m}, \tag{5.3}$$

and the mechanical power dissipated in the resonator is

$$P_m = \frac{R_m|I_m|^2}{2} = \frac{|I_m|^2}{2\omega_m Q C_m}. \tag{5.4}$$

Since the quality factor is large, the angular frequency ω is always very close to the motional resonance frequency ω_m. It is thus very useful to define a relative amount of frequency pulling (by the circuit)

$$p \triangleq \frac{\omega-\omega_m}{\omega_m} \quad \text{with} \quad |p \ll 1|. \tag{5.5}$$

The motional impedance is then given by

$$Z_m = R_m + j\omega L_m + \frac{1}{j\omega C_m} \cong R_m + j\frac{2p}{\omega C_m}, \tag{5.6}$$

where $\omega \cong \omega_m$ is considered to be constant.

The resonator is usually linear, but the circuit must be nonlinear in some way to limit the amplitude of oscillation. As a consequence, the voltage V across the resonator might be distorted. A very important assumption can be made that

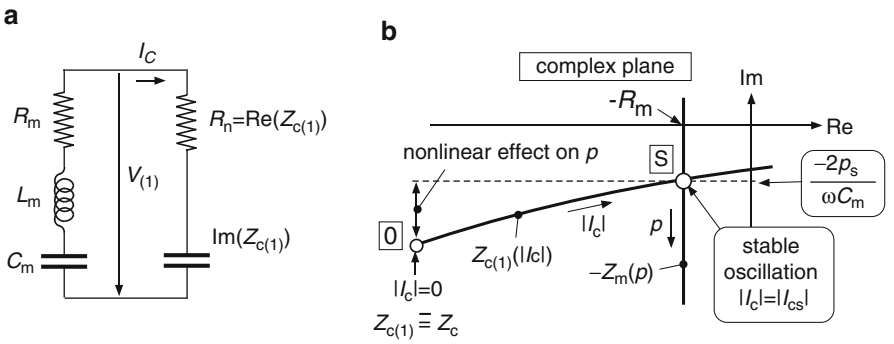

Fig. 5.2 Stable oscillation

drastically simplifies the analysis of the oscillator: since the quality factor Q is high or very high, *the motional current I_m flowing through the motional branch of the resonator is always sinusoidal*, even if the voltage V across it is strongly distorted.

Therefore, the best way to analyze the oscillator is to split it conceptually into the motional impedance Z_m, and a circuit impedance that includes the electrical component C_{12}, C_{10}, and C_{20} of the resonator, as pictured in Fig. 5.1b [2–4]. Since the current $I_c = -I_m$ flowing in the nonlinear circuit is sinusoidal, no energy can be exchanged at harmonic frequencies, and the nonlinear circuit can be characterized by its impedance for the fundamental component defined by

$$Z_{c(1)} \triangleq \frac{V_{(1)}}{I_c}, \tag{5.7}$$

where I_c is the complex value of sinusoidal current and $V_{(1)}$ the corresponding *complex value of the fundamental component of voltage V*. Because the circuit is nonlinear, $Z_{c(1)}$ is a function of the current amplitude $|I_c|$. Since the circuit contains no high-Q component, its impedance is not sensitive to small frequency variations; thus, $Z_{c(1)}$ may depend on ω but not on p.

With this definition of $Z_{c(1)}$, the condition for stable oscillation can be simply expressed by

$$Z_{c(1)}(|I_c|) = -Z_m(p). \tag{5.8}$$

This equation is explained by Fig. 5.2.

Part (a) of the figure shows the motional impedance connected to the circuit impedance $Z_{c(1)}$. The latter is composed of a real part which is a negative resistance R_n and an imaginary part supposed to be capacitive in this example. A possible locus of $Z_{c(1)}(|I_c|)$ is shown in the complex plane of part (b) of the figure. It starts at point O with the value of the small-signal impedance Z_c for very small values of $|I_c|$. The locus of $-Z_m(p)$ is represented in the same plane. According to (5.6), this locus is a straight line with constant real value $-R_m$.

In absence of oscillation (point O), the negative resistance is larger than the motional resistance, and oscillation starts to grow with a time constant

$$\tau = \frac{2L_m}{|R_n| - R_m}. \tag{5.9}$$

As soon as the amplitude $|I_c|$ of the sinusoidal current is sufficient to distort the voltage, the point operation moves to the right along the locus. According to (5.9), the time constant of growth increases until it becomes infinite at intersection point S with $-Z_m(p)$. Stable oscillation is thus reached with an amplitude $|I_{cs}|$. According to (5.6) and (5.8), the amount of frequency pulling at stable oscillation is

$$p_s = -\frac{\omega C_m \mathrm{Im}(Z_{c(1)})|_S}{2}, \tag{5.10}$$

and the variation of $\mathrm{Im}(Z_{c(1)})$ from point O to point S is proportional to the effect of nonlinearities on p.

The *critical condition for oscillation* occurs if points S and O coincide, that is, for

$$Z_c = -Z_m(p), \tag{5.11}$$

thus for

$$R_{n0} \triangleq \mathrm{Re}(Z_c) = -R_m, \tag{5.12}$$

and

$$\mathrm{Im}(Z_c) = -\frac{2p_c}{\omega C_m}, \tag{5.13}$$

where p_c is the pulling at the critical point. If the circuit is designed to minimize the nonlinear effects on the frequency, p_c becomes a very good approximation of the pulling in stable oscillation.

The linear impedance Z_c depends on the bias current applied to the circuit. Nonlinear effects can be minimized by reducing this current when the amplitude increases. With such an amplitude control loop, stable oscillation occurs very close to the critical condition, thus $Z_{c(1)}|_S \cong Z_c$.

5.2.2 *Phase Noise*

5.2.2.1 Introduction

The long-term frequency stability of the resonator itself may be degraded by the circuit. This is because the frequency pulling p_s depends on the circuit and may change with temperature, supply voltage, or drift of component values. Hence, the degradation of stability due to the circuit can be reduced by reducing p_s.

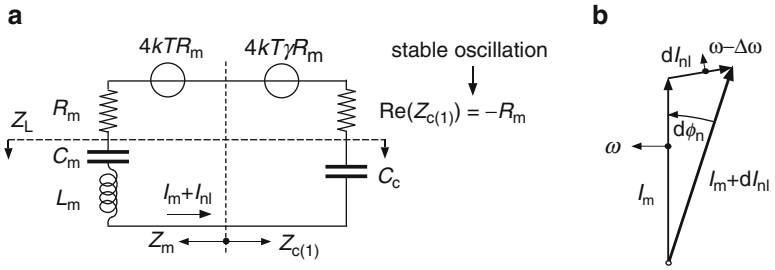

Fig. 5.3 Calculation of the basic noise: (**a**) equivalent circuit with noise sources; (**b**) phasors

Noise produces a short-term frequency instability called phase noise. Phase noise has important implications when the oscillator is used to generate the carrier frequency in RF applications.

5.2.2.2 Linear Analysis

The basic source of phase noise can be analyzed by the circuit shown in Fig. 5.3a. A thermal noise voltage of power spectral density $4kTR_m$ is associated with the motional loss resistance R_m of the resonator. At stable oscillation, the real part of the circuit impedance $Z_{c(1)}$ is equal to $-R_m$ and is associated with a noise voltage of spectral density $4kTR_m \cdot \gamma$. The factor γ is called the excess noise factor and depends on the circuit and on its bias current.

For noise frequencies ω_n close to frequency of stable oscillation ω_s, the impedance Z_L that loads the total noise voltage source of spectral density $4kTR_m(1+\gamma)$ can be expressed as

$$Z_L = 2jQR_m \frac{\Delta\omega}{\omega}, \tag{5.14}$$

where

$$\Delta\omega \triangleq \omega_n - \omega_s. \tag{5.15}$$

The spectral density of the noise current I_{nl} flowing in the loop is thus

$$S_{I_n^2} \triangleq \frac{\mathrm{d}\overline{I_{nl}^2}}{\mathrm{d}f} = \frac{4kT(1+\gamma)R_m}{|Z_L|^2} = \frac{(1+\gamma)kT}{Q^2R_m} \cdot \left(\frac{\omega}{\Delta\omega}\right)^2. \tag{5.16}$$

This noise current is added to the motional current I_m. For an elementary bandwidth $\mathrm{d}f$ at angular frequency ω, the corresponding complex phasors are illustrated in Fig. 5.3b at a given instant. The length of the phasor $\mathrm{d}I_{nl}$ is a random value of variance $S_{I_n^2}\mathrm{d}f$. Since the noise current is small compared to the motional current, the power spectrum of the phase noise ϕ_n can be expressed as

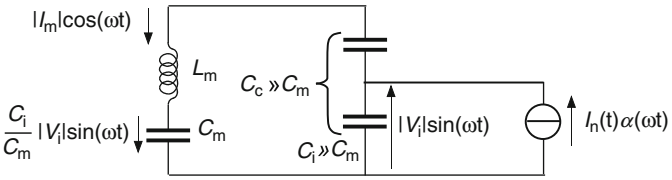

Fig. 5.4 Equivalent circuit for nonlinear analysis of phase noise

$$S_{\phi_n^2} \triangleq \frac{\mathrm{d}\overline{\phi_n^2}}{\mathrm{d}f} = \frac{1}{2}\frac{S_{I_n^2}}{(|I_m|/\sqrt{2})^2} = \frac{(1+\gamma)kT\omega^2}{Q^2R_m|I_m|^2(\Delta\omega)^2}, \tag{5.17}$$

where the factor 1/2 comes from the fact that the other half of the noise is amplitude noise. The dependency of phase noise on the power P_m dissipated in the resonator or the energy of mechanical oscillation is obtained by introducing (5.3) and (5.4) in (5.17):

$$S_{\phi_n^2} = \frac{(1+\gamma)kT\omega^2}{2Q^2P_m(\Delta\omega)^2} = \frac{(1+\gamma)kT\omega}{2QE_m(\Delta\omega)^2}. \tag{5.18}$$

5.2.2.3 Nonlinear Time Variant Circuit

When the amplitude increases, the oscillator circuit becomes nonlinear, and the noise produced by the transistors becomes cyclostationary [5]. Phase noise can then be analyzed by using the impulse sensitivity function (ISF) introduced by Hajimiri and Lee [6], by means of the equivalent circuit for stable oscillation shown in Fig. 5.4.

A cyclostationary noise source current is injected in a capacitor C_i that is part of the circuit capacitance C_c in series with the motional capacitance C_m of the resonator. This noise source is composed of a stationary noise current $I_n(t)$ multiplied by a modulation function $\alpha(\omega t)$ synchronous with the voltage across C_i.

According to [6], the phase noise spectrum density for a white noise current source $I_n(t)$ is given by

$$S_{\phi_n^2} = \frac{\overline{\Gamma^2}\cdot S_{I_n^2}}{2(C_i|V_i|)^2\Delta\omega^2}\cdot\left(\frac{C_m}{C_i}\right)^2. \tag{5.19}$$

The product $C_i|V_i|$ is the maximum charge in capacitors C_i and C_m. The factor C_m/C_i corresponds to the fraction of the total voltage (state variable) that appears across $C_i \gg C_m$. The variable $\Delta\omega$ is the offset of the noise frequency ω_n with respect to the frequency of stable oscillation defined by (5.15).

The periodic function Γ is the *effective impulse sensitivity function*. Since the quality factor Q is very large, the motional current is perfectly sinusoidal. The same is true for the voltage across the motional capacitance C_m. To obtain an analytic

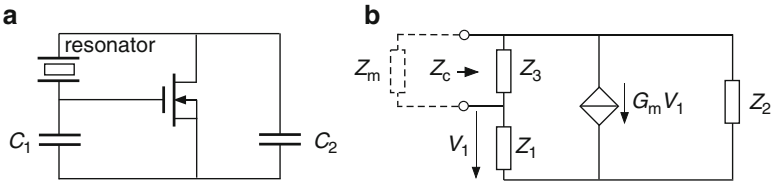

Fig. 5.5 (**a**) Basic Pierce oscillator without biasing circuitry; (**b**) small-signal equivalent circuit

solution for phase noise, we will assume that the *voltage across C_i is sinusoidal* as well. This is an *approximation*, since this voltage is slightly distorted by the current delivered by the active device to compensate the losses in the resonator. With this approximation, the effective ISF for a sinusoidal voltage $|V_i| \sin \phi$ across C_i derived in [6] becomes simply

$$\Gamma = \alpha(\phi) \cos \phi, \tag{5.20}$$

where $\phi = \omega t$. As shown by (5.19), the phase noise spectrum due to white noise sources depends on *average value of the square* of the ISF. For a 1/f flicker noise current of spectral density K_{fi}/ω_n, the phase noise spectrum depends on the *square of the average value* of the ISF, according to

$$S_{\phi_n^2} = \frac{\overline{\Gamma}^2 \cdot K_{fi}}{(C_i|V_i|)^2 \Delta \omega^3} \cdot \left(\frac{C_m}{C_i}\right)^2. \tag{5.21}$$

The effective ISF is still defined by (5.20), but it may be different from that for white noise.

5.3 The Pierce Oscillator

5.3.1 Basic Circuit and Linear Analysis

The most simple quartz or MEMs oscillator is the Pierce circuit illustrated in Fig. 5.5a [1].

The circuit itself uses a single 3-terminal active transistor (source connected to the local substrate) and two functional capacitors C_1 and C_2. The resonator is connected between the gate and the drain. It is very similar to the Clapp oscillator [7], except for the presence of the unavoidable capacitance in parallel with the series resonator, which has important implications in the circuit behavior.

The small-signal equivalent circuit of this oscillator is shown in Fig. 5.5b. Each impedance Z_i is defined as

$$Z_i \triangleq \frac{1}{j\omega C_i + G_i}. \tag{5.22}$$

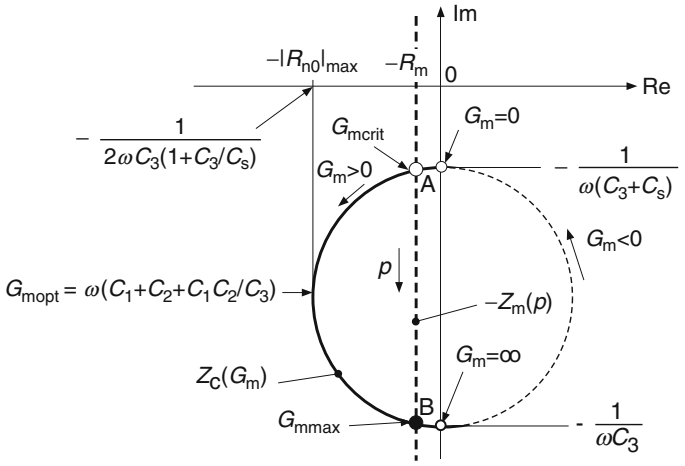

Fig. 5.6 Loci of $Z_c(G_m)$ and $-Z_m(p)$ for the lossless Pierce circuit

The capacitor C_3 includes C_{12} of the resonator and some parasitic capacitors. The functional capacitors C_1 and C_2 include C_{10} and C_{20} and parasitic capacitors. The transistor is saturated with a transconductance G_m. The loss conductances G_i are due to the bias circuitry and other possible losses.

The small-signal impedance of the circuit is

$$Z_c = \frac{Z_1Z_3 + Z_2Z_3 + G_mZ_1Z_2Z_3}{Z_1 + Z_2 + Z_3 + G_mZ_1Z_2}. \tag{5.23}$$

It is a *bilinear function of* G_m. Therefore, the locus of $Z_c(G_m)$ in the complex plane is a *circle*. If the circuit is lossless, with $G_1 = G_2 = G_3 = 0$, then the circle is centered on the imaginary axis, as depicted in Fig. 5.6.

For $G_m = 0$, the circuit impedance is imaginary and has the value

$$Z_c|_{G_m=0} = -\frac{j}{\omega(C_3 + C_s)}, \tag{5.24}$$

where C_s is the series combination of C_1 and C_2:

$$C_s = \frac{C_1C_2}{C_1 + C_2}. \tag{5.25}$$

It is also imaginary for G_m infinite:

$$Z_c|_{G_m=\infty} = -\frac{j}{\omega C_3}. \tag{5.26}$$

The maximum value of small-signal negative resistance is given by the radius of the circle:

$$|R_{n0}|_{max} = \frac{1}{2\omega C_3(1+C_3/C_s)}. \tag{5.27}$$

It is obtained for a value of G_m called G_{mopt} with the value given in the figure.

The locus of $-Z_m(p)$ is also reported on the figure. According to (5.6), it is a vertical line at position $-R_m$. It intersects $Z_c(G_m)$ at points A and B. Point A corresponds to the critical condition for oscillation, which is obtained for the critical value of transconductance G_{mcrit}. For $G_m > G_{mcrit}$, the oscillation grows until the amplitude reaches a level for which the circuit cannot be considered linear any longer. For $G_m = G_{mopt}$, the excess of negative resistance $|R_{n0}| - R_m$ is maximum, and the time constant of growth τ given by (5.9) is minimum. A further increase of G_m reduces $|R_{n0}| - R_m$ until it comes to zero for $G_m = G_{mmmax}$ at point B. No oscillation can start for $G_m > G_{mmmax}$.

Of course, oscillation is only possible if $|R_{n0}|_{max} > R_m$. By combining (5.27) with (5.1), this condition can be expressed as

$$M \triangleq \frac{QC_m}{C_3} > 2(1+C_3/C_s) > 2, \tag{5.28}$$

where M is the figure of merit of the resonator. This condition expresses the fact that the overall impedance between drain and gate must be inductive.

According to (5.13), the pulling is proportional to the imaginary part of Z_c. Thus, in order to avoid a dependency of the frequency on R_m (thus on the quality factor Q), Fig. 5.6 shows that R_m should be much smaller than $|R_{n0}|_{max}$ (condition (5.28) fulfilled with a large margin). If this is the case, then from (5.13),

$$p_c = \frac{C_m}{2(C_3+C_s)}, \tag{5.29}$$

and it can be shown from (5.12) and (5.23) that

$$G_{mcrit} = \omega^2 R_m C_1 C_2 (1+C_3/C_s)^2 = \frac{\omega}{QC_m} C_1 C_2 (1+C_3/C_s)^2, \tag{5.30}$$

whereas the maximum transconductance for oscillation is

$$G_{mmax} = \frac{C_1 C_2}{R_m C_3^2} = \frac{\omega C_m C_1 C_2 Q}{C_3^2}. \tag{5.31}$$

The dependency of G_{mcrit} on p_c can be expressed by introducing (5.29) in (5.30), which yields

$$G_{mcrit} = \frac{\omega C_m}{Qp_c^2} \cdot \frac{(C_1+C_2)^2}{4C_1C_2} \underbrace{=}_{\text{for } C_1=C_2} \frac{\omega C_m}{Qp_c^2}. \tag{5.32}$$

The critical transconductance for a given amount of pulling is minimum for $C_1 = C_2$.

If the loss conductances are not negligible, the circular locus is moved to the right. It can be shown that the resulting increase of critical transconductance can be approximated by

$$\Delta G_{mcrit} \cong \frac{G_1 C_2^2 + G_2 C_1^2 + G_3 (C_1 + C_2)^2}{C_1 C_2} \underbrace{=}_{\text{for } C_1 = C_2} G_1 + G_2 + 4G_3. \tag{5.33}$$

5.3.2 Amplitude of Oscillation

If the bias current I_0 of the transistor is such that $G_{mcrit} < G_m < G_{mmax}$, the amplitude of oscillation grows and is first limited by the nonlinear transfer characteristics of the transistor. The gate voltage V_1 can be shown to remain almost sinusoidal, but the drain current is distorted. Since its DC value must remain constant at value I_0, the fundamental component of drain current $I_{D(1)}$ is reduced. The equivalent transconductance for the fundamental, defined as

$$G_{m(1)} \triangleq |I_{D(1)}| / |V_1|, \tag{5.34}$$

is reduced by the increasing amplitude until stable oscillation is reached for $G_{m(1)} = G_{mcrit}$.

If the transistor is in weak inversion (with exponential transfer characteristics) and remains saturated, the bias current I_0 needed to obtain an amplitude $|V_1|$ at the gate is given by

$$\frac{I_0}{I_{0critmin}} = \frac{\mathrm{I_{B0}}(v_1)}{2\mathrm{I_{B1}}(v_1)} v_1, \tag{5.35}$$

where

$$I_{0critmin} = n U_T G_{mcrit}, \tag{5.36}$$

is the current needed in weak inversion to produce the critical transconductance. $U_T = kT/q$ is the thermodynamic voltage, and n is the slope factor (ranging between 1.2 and 1.5). $\mathrm{I_{B0}}$ and $\mathrm{I_{B1}}$ are modified Bessel functions of order zero and one, and

$$v_1 = \frac{|V_1|}{n U_T} \tag{5.37}$$

is the normalized gate voltage amplitude. This result is plotted in thick line in Fig. 5.7. It is close to the strait dotted line, which would be obtained by Dirac drain current pulses having the same mean value I_0.

In strong (and medium) inversion, the $|V_1|(I_0)$ relationship depends on the inversion coefficient IC_0 at the critical current, defined as

$$IC_0 = I_{0crit} / I_{spec}, \tag{5.38}$$

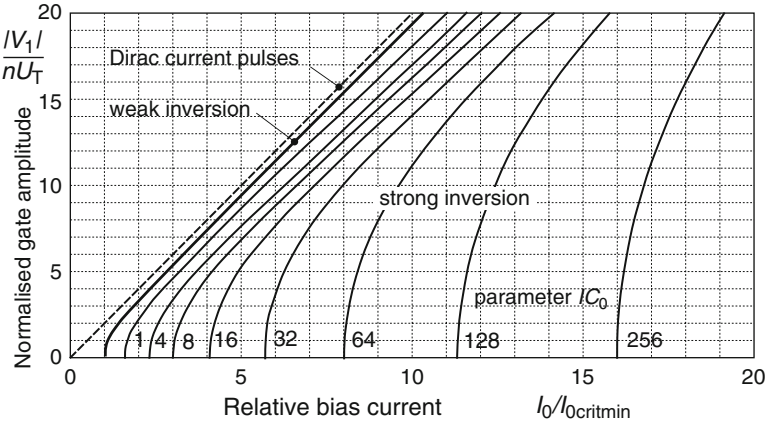

Fig. 5.7 Normalized gate amplitude; the bias current I_0 is normalized to its minimum critical value that is reached in weak inversion; $IC_0 = I_{0crit}/I_{spec}$ is the inversion coefficient at the critical condition for oscillation ($|V_1| = 0$)

where

$$I_{spec} = 2n\beta U_T^2 = 2n\mu C_{ox} U_T^2 W/L \tag{5.39}$$

is the specific current of the transistor that depends on the width-to-length ratio W/L. No simple analytic expression exists if the drain current is cut off during the negative peaks of gate voltage. Curves for various values of IC_0 have been obtained by numerical computation on the EKV model [8,9] and are also plotted in the figure.

The nonlinear effects due to the transfer characteristics are essentially equivalent to a reduction of the small-signal transconductance G_m (this is exactly true if Z_3 is very large, so that V_1 remains perfectly sinusoidal). Thus, the locus of $Z_{c(1)}(I_c)$ follows the circular locus of $Z_c(G_m)$. For a given value of R_m, $p_c \cong p_s$ and is independent of the amplitude.

This would not be the case for stronger nonlinear effects such as those due to the desaturation of the active transistor or of the transistor delivering the bias current. In order to avoid those effects, the amplitude should be limited. This is best obtained by using an amplitude control loop as described in [2, 10] or [11].

5.3.3 Phase Noise

5.3.3.1 Introduction

For the evaluation of phase noise, the circuit impedance Z_3 between drain and gate can be neglected in a first approximation. The noise produced by the transistor or by its biasing current source is then simply flowing through capacitor C_2.

5.3.3.2 Linear Calculation

For a noise frequency ω_n close to the oscillation frequency ω, a noise current of spectral density $S_{I_n^2}$ produces a noise voltage of spectral density

$$S_{V_n^2} = \frac{S_{I_n^2}}{(\omega C_2)^2} \tag{5.40}$$

in series with Z_m and Z_c. According to Fig. 5.3, the noise excess factor for a very small amplitude is thus

$$\gamma_0 = \frac{S_{V_n^2}}{4kTR_m} = \frac{S_{I_n^2}}{4kTR_m(\omega C_2)^2}. \tag{5.41}$$

The spectral density of channel noise current of a saturated transistor can be expressed as [9]

$$S_{I_{nD}^2} = 4kTn\gamma_t G_m, \tag{5.42}$$

where n is the slope factor (ranging from 1.2 to 1.5) and γ_t is the noise excess factor of the transistor (normally 1/2 in weak inversion and 2/3 in strong inversion).

Introducing this expression in (5.41) with $G_m = G_{mcrit}$ and using the expression (5.30) of G_{mcrit} to replace R_m gives the phase noise excess factor for small amplitudes due to the transistor:

$$\gamma_0 = n\gamma_t C_1/C_2. \tag{5.43}$$

If G_{mcrit} is increased by losses, then γ_0 is increased proportionally.

5.3.3.3 Phase Noise of the Nonlinear Time Variant Circuit

As discussed before, the results obtained by the linear analysis are no longer valid when the amplitude increases. The circuit becomes nonlinear and the noise becomes cyclostationary, with a noise modulation function α (defined in Fig. 5.4) that depends on the mode of operation of the transistor.

If the transistor operates in *weak inversion*, calculations of the spectrum of the *white* cyclostationary noise due to channel noise give

$$\alpha^2 S_{I_n^2} = \underbrace{nkTG_{mcrit}\frac{v_1}{\mathrm{I_{B1}}(v_1)}}_{S_{I_n^2}} \underbrace{\mathrm{e}^{v_1\sin(\phi+\Delta\phi)}}_{\alpha^2}, \tag{5.44}$$

where the first part is the spectrum $S_{I_n^2}$ of a fictitious white noise current modulated by α, and $\Delta\phi$ is the small phase shift between the gate and drain voltages. According to (5.20), the squared RMS value of the effective impulse sensitivity function (ISF) is

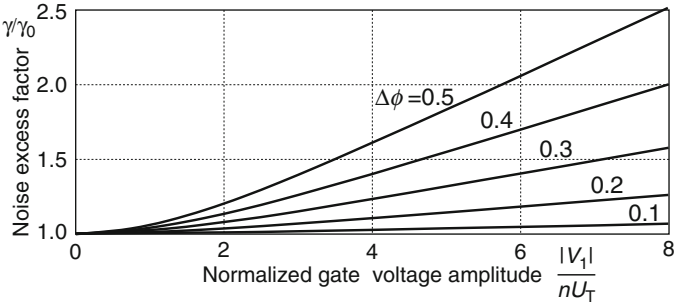

Fig. 5.8 Variation of the phase noise excess factor γ in weak inversion

$$\overline{\Gamma^2(v_1)} = \overline{\cos^2\phi\alpha^2} = \frac{1}{2\pi}\int_0^{2\pi} \cos^2\phi \cdot e^{v_1\sin(\phi+\Delta\phi)} d\phi \quad \text{(weak inversion)}. \tag{5.45}$$

The phase noise spectrum is then obtained by introducing the above values of $S_{I_n^2}$ and $\overline{\Gamma^2}$ in (5.19), which results in

$$S_{\phi_n^2} = S_{\phi_n^2 0}\frac{v_1}{I_{B1}(v_1)}\overline{\Gamma^2(v_1)}, \tag{5.46}$$

where

$$S_{\phi_n^2 0} = \frac{nC_1}{2C_2}\cdot\frac{kT}{Q^2|I_m|^2 R_m}\cdot\frac{\omega^2}{\Delta\omega^2}. \tag{5.47}$$

According to (5.17), this corresponds to an excess noise factor γ_0 given by (5.43) that was obtained by the linear analysis, since $\gamma_t = 1/2$ in weak inversion. For larger amplitudes, γ grows proportionally to the noise spectrum contribution:

$$\gamma = \gamma_0\frac{v_1}{I_{B1}(v_1)}\overline{\Gamma^2(v_1)}. \tag{5.48}$$

This variation is plotted in Fig. 5.8 for several values of phase shift $\Delta\phi$.

This figure shows that, in weak inversion, the phase noise excess factor γ remains constant and equal to its small amplitude value γ_0 if there is no phase shift $\Delta\phi$. Otherwise, it increases slowly with the amplitude.

Calculations for *strict strong inversion* (i.e., for $|V_1| \le 2nU_T\sqrt{IC_0}$ so that the current is not cut off in the negative peaks of gate voltage) show that γ remains constant independently of the phase shift $\Delta\phi$.

The *flicker noise* of a MOS transistor can be *approximated* by a noise voltage V_{nG} at the gate with a spectrum

$$S_{V_{nG}^2} = K_f/\omega_n \tag{5.49}$$

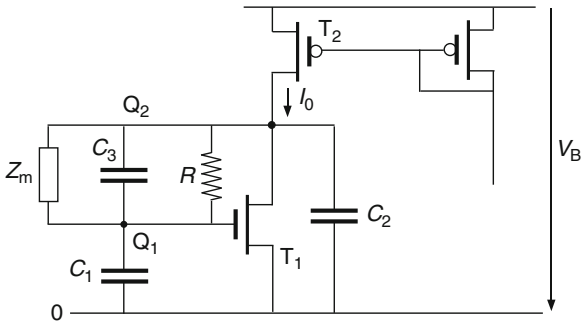

Fig. 5.9 Grounded-source implementation of the Pierce oscillator

independent of the bias current. The resulting cyclostationary noise spectrum obtained for *weak inversion* is

$$\alpha^2 S_{I_n^2} = \frac{K_f G_m^2}{\omega_n} = \underbrace{\frac{K_f G_{mcrit}^2}{\omega_n}\left(\frac{v_1}{2\mathrm{I_{B1}}(v_1)}\right)^2}_{K_{fi}/\omega_n}\underbrace{\left(\mathrm{e}^{v_1\sin(\phi+\Delta\phi)}\right)^2}_{\alpha^2}, \tag{5.50}$$

where the first underbraced part is the spectrum density K_{fi}/ω_n of a fictitious stationary flicker noise current source modulated by α. The average value of the ISF is thus

$$\overline{\Gamma} = \frac{1}{2\pi}\int_0^{2\pi} \cos\phi \mathrm{e}^{v_1\sin\phi+\Delta\phi}\mathrm{d}\phi = \sin\Delta\phi\cdot \mathrm{I_{B1}}(v_1). \tag{5.51}$$

Introducing the above values of K_{fi} and $\overline{\Gamma}$ in (5.21) and using the expression (5.30) of G_{mcrit} (with $C_3 = 0$) yields finally

$$S_{\phi_n^2} = \left(\frac{\omega}{2QnU_T}\right)^2 \cdot \frac{K_f}{\Delta\omega^3}\cdot \sin^2\Delta\phi, \tag{5.52}$$

which is independent of the amplitude, but depends on the phase shift $\Delta\phi$ between gate and drain voltages.

Similar calculations for *strong inversion* show that the value of $S_{\phi_n^2}$ given by (5.52) is reduced by a factor $4IC_0$.

5.3.4 Practical Implementations

The best way of implementing the Pierce oscillator is to ground the source of the transistor, as illustrated in Fig. 5.9.

The two functional capacitors are then grounded, but not the resonator, which requires two special connecting pins at nodes $\mathrm{Q_1}$ and $\mathrm{Q_2}$. The active transistor $\mathrm{T_1}$ is biased by a current I_0 delivered by the p-channel current source $\mathrm{T_2}$ and by resistor R that forces the DC components of gate and drain voltages to be equal. The value of

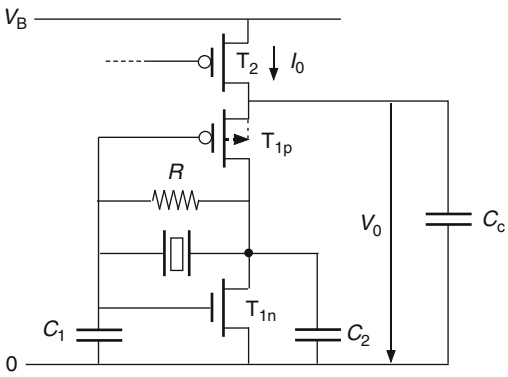

Fig. 5.10 CMOS implementation of the Pierce oscillator

R must be large to minimize the loss conductance G_3. If resistors are not available, R may be implemented by means of a transistor adequately biased [2].

To avoid additional losses and nonlinearities, the gate amplitude must be limited to avoid desaturation of the transistors in the peaks of oscillation. The DC component at the drains depends on the threshold voltage of T_1 and on the amplitude of oscillation. It may be better centers by flowing a very small current through R.

Replacing the single active transistor by a complementary pair reduces the bias current needed to obtain the necessary transconductance. However, a realization based on the simple CMOS gate is not recommended because of its many disadvantages. The features of a CMOS pair are best exploited by the current-biased configuration depicted in Fig. 5.10 [12].

The CMOS inverter $T_{1n} - T_{1p}$ is biased by a current I_0 delivered by T_2. It produces a transconductance $G_m = G_{m1n} + G_{m1p}$. A coupling capacitance C_c is needed to ground the source of the p-channel transistor T_{1p} for the AC signal of oscillation. Its admittance must be much larger than the source transconductance nG_{m1p} of T_{1p}.

The DC voltages at the drains and V_0 across the CMOS inverter depend on the bias current, on the threshold voltages of the transistors, and on the amplitude of oscillation. The latter must be limited to avoid desaturation of T_{1n} and/or T_{1p} in the peaks of oscillation. To allow of a larger amplitude, the DC voltages can be increased by separately biasing the two complementary transistors [2, 12].

The transistor T_{1p} can be put in a separate well connected to its source, as indicated in dotted line in the figure. This provides a better isolation of the inverter from variations of the supply voltage V_B.

The basic Pierce oscillator can also be implemented by grounding the drain of the active transistor, as shown in Fig. 5.11. It has the advantage of requiring only one special connection pin for the resonator, since the other side is grounded for AC currents. But the functional capacitor C_1 is now floating.

The main drawback is that the unwanted capacitor C_3 is increased since it includes one of the parasitic capacitors C_{10} or C_{20} defined in Fig. 5.1a, plus the

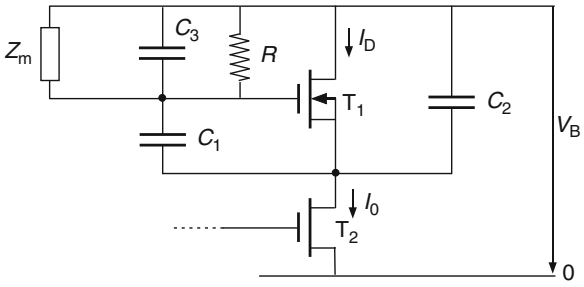

Fig. 5.11 Grounded-drain implementation of the Pierce oscillator

capacitance of the interconnection with respect to ground. The radius of the circular locus of Fig. 5.6 may thus be considerably reduced, thereby reducing the margin of available negative resistance.

The active transistor T_1 must be put in a separate well connected to the source as shown in the figure. Indeed, if this transistor is in the common substrate connected to the negative supply rail, the current is also modulated by the source, with a transconductance nG_m. Then, it can then be shown that

$$\mathrm{Im}(Z_c)|_{G_m=\infty} = \frac{-1}{\omega\left(C_3 + \frac{n-1}{n}C_1\right)} \quad \text{instead of} \quad \frac{-1}{\omega C_3}. \tag{5.53}$$

Since C_1 is usually much larger than C_3, the radius of the circular locus of Fig. 5.6 is dramatically reduced.

Grounding the gate of the active transistor, as is often done for the Colpitts oscillator, is not convenient since there is no DC path through the resonator.

5.4 Parallel-Resonance Oscillator

5.4.1 Basic Circuit and Linear Analysis

Because the resonator includes the capacitor C_0 in parallel with the motional impedance, its global impedance has a maximum that corresponds to a parallel resonance (also called antiresonance). The standard symmetrical circuit for a parallel L, C, G resonant circuit is depicted in Fig. 5.12a.

The DC path across the inductance prevents this cross-coupled circuit from becoming a bistable flip-flop. But this DC path does not exist for the electromechanical resonator. Thus, the DC coupling of the sources must be replaced by an AC coupling through capacitor C_S, as shown in part (b) of the figure. It can be shown that the circuit without Z_m remains stable (no parasitic oscillation) only if

$$C_S < nC_D, \tag{5.54}$$

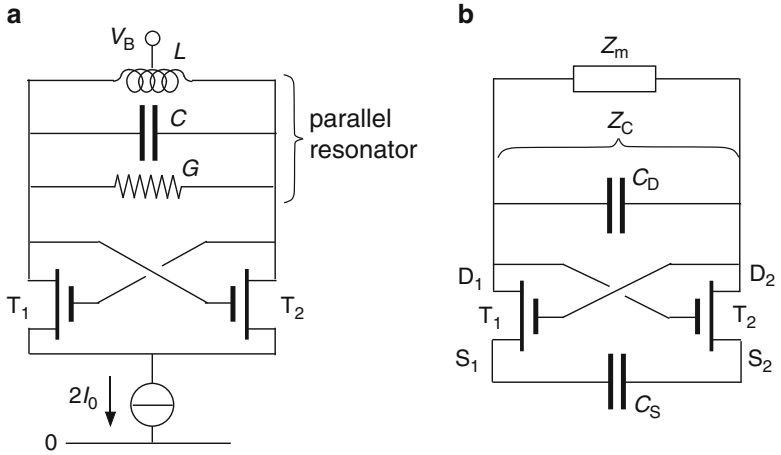

Fig. 5.12 Parallel-resonance oscillators: (**a**) standard $L-C-G$ oscillator; (**b**) modified circuit for the electromechanical resonator (biasing circuitry omitted)

where C_D includes the parallel capacitance C_0 of the resonator and n is the slope factor of the transistors (assumed to be in a common substrate connected to the negative supply rail).

Although this oscillator is basically exploiting the parallel resonance of the resonator, it can still be analyzed by spitting it into the motional impedance Z_m and the circuit impedance Z_c. The latter is then

$$Z_c = \frac{Z_D(2+nG_mZ_S)}{2+nG_mZ_S-G_mZ_D}, \tag{5.55}$$

where

$$Z_D \triangleq \frac{1}{j\omega C_D+G_D} \quad \text{and} \quad Z_S \triangleq \frac{1}{j\omega C_S+G_S} \tag{5.56}$$

include possible loss conductances G_D and G_S. The impedance Z_c is once more a bilinear function of G_m. The corresponding circular locus is shown in Fig. 5.13 for the lossless case, and for three different ratios C_S/C_D.

At the limit of condition (5.54), with $C_S = nC_D$, the locus is a horizontal straight line (circle of infinite radius) at $-1/\omega C_D$. If the condition is not fulfilled, the circular locus is located above this limit line. The circle for $C_S < nC_D$ is located underneath this line, and Z_c remains capacitive. The maximum negative resistance is given by the radius of the circle:

$$|R_{n0max}| = \frac{C_S}{2\omega C_D(nC_D-C_S)}, \tag{5.57}$$

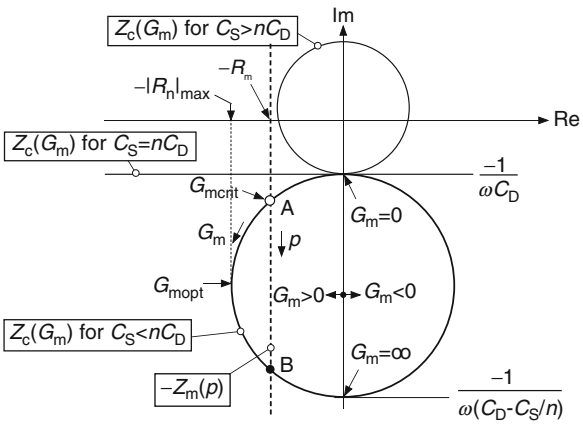

Fig. 5.13 Loci of $Z_c(G_m)$ for the parallel-resonance circuit of Fig. 5.12b; a possible locus of $-Z_m(p)$ is also represented in *dotted line*

and is reached for $G_m = G_{mopt}$ given by

$$G_{mopt} = \frac{2\omega C_D C_S}{nC_D - C_S}. \tag{5.58}$$

If $R_m < |R_{n0}|_{max}$, the locus of $-Z_m(p)$ intersects the circle at points A and B. Point A is stable and gives the critical condition for oscillation. If $R_m \ll |R_{n0}|_{max}$, then the imaginary part of Z_c is $-1/\omega C_D$. Thus, according to (5.13), the pulling is

$$p_c = \frac{C_m}{2C_D}. \tag{5.59}$$

It can be shown that, again for $R_m \ll |R_{n0}|_{max}$, the critical transconductance is

$$G_{mcrit} = \frac{2\omega C_D^2}{QC_m} = 2\omega^2 C_D^2 R_m = \frac{\omega C_m}{2Qp_c^2}. \tag{5.60}$$

A small loss conductance $G_D \ll \omega C_D$ increases the critical transconductance by about $2G_D$. It also increases the effective value of C_D. Thus, the margin of stability with respect to the condition (5.54) is increased, and the pulling is slightly reduced.

A loss resistance R_S in series with C_S has no effect on the pulling, but it increases the inverse of the source transconductance of each transistor by $R_S/2$. The critical transconductance G_{mcrit} is therefore increased.

5.4.2 *Amplitude of Oscillation*

The most direct way to implement the parallel-resonance oscillator of Fig. 5.12b is depicted in Fig. 5.14.

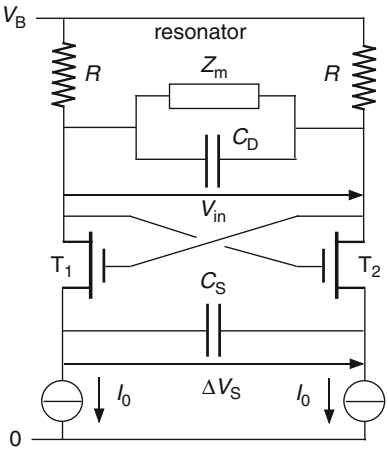

Fig. 5.14 Simple biasing of the parallel-resonance oscillator

Each transistor is biased by a current source I_0, and load resistors R fix the common mode drain and gate voltage at $V_B - RI_0$. Their value must be sufficiently large to limit the corresponding loss conductance $G_D = 1/2R$.

The circuit can be seen as a differential pair biased by $2I_0$ and degenerated by the impedance of C_s. It can be shown that this degeneration is negligible if

$$\frac{QC_m}{C_D} \gg \frac{nC_D}{C_S}, \tag{5.61}$$

which is usually the case.

Moreover, the parallel resonator made of Z_m and C_D has a high Q. Thus, the voltage V_{in} across it remains sinusoidal even if the drain currents are distorted. This voltage is the input voltage of the differential pair.

A transconductance for the fundamental component $G_{m(1)}(V_{in})$ can be calculated by using the analytic expression of the transfer characteristics of the differential pair $(I_{D1} - I_{D2})(V_{in})$. As was done for the Pierce oscillator, the current-amplitude relationship is then obtained by equating $G_{m(1)}$ with G_{mcrit}.

The result obtained for transistors operated in *weak inversion* is

$$\frac{I_0}{I_{0crit}} = \frac{I_0}{nU_T G_{mcrit}} = \frac{v_{in}}{2\mathrm{f_w}(v_{in})}, \tag{5.62}$$

where

$$v_{in} \triangleq \frac{V_{in}}{nU_T} \quad \text{and} \quad \mathrm{f_w}(v_{in}) = \frac{1}{\pi}\int_0^{2\pi} \tanh\left(\frac{v_{in}}{2}\sin\phi\right)\sin\phi\,\mathrm{d}\phi. \tag{5.63}$$

It is plotted in Fig. 5.15. Results obtained for several levels of strict strong inversion are also reported in the same plot. Here, the level of inversion is

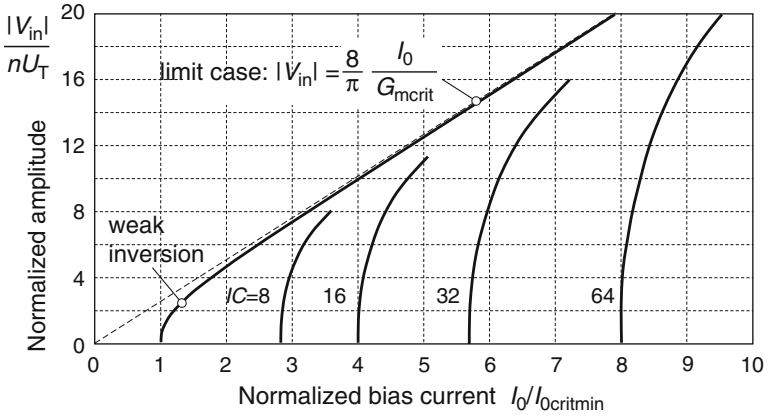

Fig. 5.15 Amplitude of oscillation of the parallel-resonance oscillator when condition (5.61) is fulfilled. The transistors are assumed to remain saturated

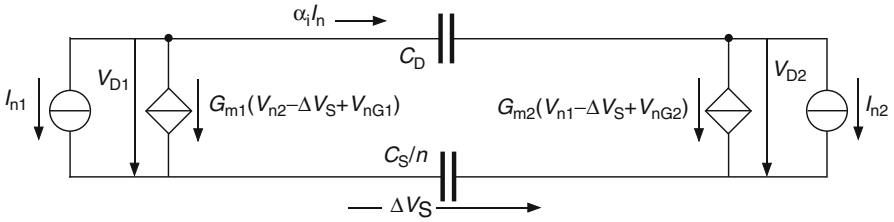

Fig. 5.16 Equivalent circuit for the calculation of noise

characterized by the inversion coefficient at stable amplitude $IC = I_0/I_{spec}$. The curves for strong inversion therefore do *not* represent the variation of amplitude for a fixed-size transistor.

An asymptotic limit

$$\frac{V_{in}}{nU_T} = \frac{8}{\pi} \cdot \frac{I_0}{G_{mcrit}} \tag{5.64}$$

is approached for $I_0/I_{0crit} \gg 1$, when the drain currents tend to a square wave shape (the total bias current $2I_0$ flows through one of the transistors).

5.4.3 Phase Noise

5.4.3.1 Introduction

The noise analysis of the basic lossless circuit of Fig. 5.12b can be carried out with the equivalent circuit of Fig. 5.16.

The source-to-source capacitance C_S is divided by n to take into account the fact that the source transconductance is n-times larger than the gate transconductance. Although the circuit is physically symmetrical, the transconductances G_{m1} and G_{m2} of the transistors vary in opposite phase along each cycle of oscillation. Therefore, their values are not equal. Furthermore, the noise sources I_{n1}, I_{n2}, V_{nG1}, and V_{nG1} are not correlated. Calculations with this circuit yield

$$\alpha_i I_n = \frac{G_{m2} I_{n1} - G_{m1} I_{n2} + G_{m1} G_{m2} (V_{nG1} + V_{nG2})}{(G_{m1} + G_{m2}) + \underbrace{G_{m1} G_{m2} (n/C_S - 1/C_D)/j\omega_n}_{\text{neglected}}} \quad . \tag{5.65}$$

Now, C_S is normally close to nC_D (to maximize the radius of the locus of $Z_c(G_m)$. Thus, for noise frequencies that are not too low, the imaginary part of the denominator *can be neglected*.

5.4.3.2 Linear Analysis

If the noise currents I_{n1} and I_{n2} are solely the white channel noise currents of the transistors, with spectral densities given by (5.42), then the spectrum of stationary noise voltage V_n across C_D is obtained from (5.65) with $G_{m1} = G_{m2} = G_{mcrit}$:

$$S_{V_n^2} = 2kTn\gamma_t \cdot \frac{G_{mcrit}}{(\omega C_D)^2} = 4kTn\gamma_t R_m. \tag{5.66}$$

Hence, referring to Fig. 5.3a, the phase noise excess factor for small amplitude is

$$\gamma_0 = \frac{S_{V_n^2}}{4kTR_m} = n\gamma_t. \tag{5.67}$$

If G_{mcrit} is increased by losses, then γ_0 is increased proportionally.

5.4.3.3 Phase Noise of the Nonlinear Time Variant Circuit

The spectrum of the cyclostationary noise current due to channel noise currents I_{n1} and I_{n2} can be calculated from (5.65) and (5.42). The result for *weak inversion* is

$$\alpha^2 S_{I_n^2} = \underbrace{2n\gamma_t kTG_{mcrit} \cdot \frac{v_{in}}{2\mathrm{f_w(v_{in})}}}_{S_{I_n^2}} \cdot \underbrace{\frac{4}{2 + \mathrm{e}^{v_{in}\sin\phi} + \mathrm{e}^{-v_{in}\sin\phi}}}_{\alpha^2}. \tag{5.68}$$

The squared RMS value of the effective ISF is thus

$$\overline{\Gamma^2} = \overline{\cos^2(\phi + \Delta\phi)\alpha^2} = \frac{1}{2\pi}\int_0^{2\pi} \frac{4\cos^2(\phi + \Delta\phi)}{2 + \mathrm{e}^{v_{in}\sin\phi} + \mathrm{e}^{-v_{in}\sin\phi}} \mathrm{d}\phi, \tag{5.69}$$

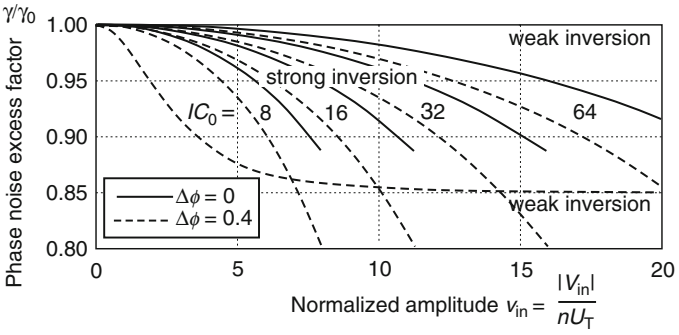

Fig. 5.17 Variation of the phase noise excess factor due to channel white noise with the drain-to-drain amplitude

where $\Delta\phi$ is the small phase shift between the gate voltages and the drain currents due to the source-to-source capacitor C_S.

The phase noise spectrum is then calculated by introducing the above values of $S_{I_n^2}$ and $\overline{\Gamma^2}$ in (5.19), with $C_i = C_D$ and $V_i = V_{in}$. The corresponding excess phase factor is finally obtained by dividing this spectrum by $4kTR_m$, according to Fig. 5.3a. The result is

$$\gamma = \gamma_0 \frac{v_{in}}{f_w(v_{in})} \overline{\Gamma^2}, \tag{5.70}$$

where γ_0 is given by (5.67). This result is plotted in Fig. 5.17 for two values of phase shift $\Delta\phi$.

Results obtained for transistors operated strictly in strong inversion (no cut off of currents) are also reported on this plot, for several values of the inversion coefficient IC_0 defined by (5.38), but approximated by $IC = I_0/I_{spec}$.

Except for weak inversion with $\Delta\phi = 0$, γ decreases slightly when the amplitude increases.

In the expression (5.65) of the cyclostationary noise current, the factor weighting V_{nG1} and V_{nG1} is symmetrical with respect to G_{m1} and G_{m2}. As a consequence, the corresponding value of $\overline{\Gamma}$ is zero, and no flicker noise is transposed around the oscillation frequency. But this is only true with the simple model of a bias-independent noise voltage at the gate, corresponding to a drain current noise spectral density proportional to G_m^2. This model is approximately valid if the inversion coefficient is close to unity [9]. If the noise current density is proportional to a different power of the transconductance, then the symmetry is lost and some flicker noise is transposed around the oscillation frequency. The resulting phase noise spectrum is proportional to $\sin^2 \Delta\phi$ and maximum at small amplitudes. It is reduced when the amplitude increases [2].

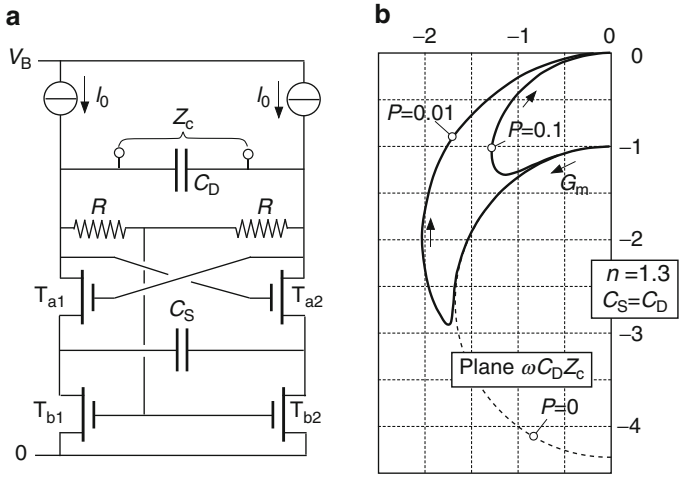

Fig. 5.18 Practical implementation of the parallel-resonance oscillator: (**a**) circuit; (**b**) locus of $Z_c(G_m)$ for two different values of P; the locus for the lossless circuit ($P = 0$) is shown in *dotted line*

5.4.4 Practical Implementations

The load resistors R in the direct implementation of Fig. 5.14 must be large to limit losses. Therefore, it is not compatible with a low supply voltage V_B. An alternative solution, better suited to low voltage, is illustrated in Fig. 5.18a.

The load resistors are replaced by current sources I_0. The common mode drain and gate voltage of T_a is controlled by a feedback loop through the biasing transistors T_b. This common mode voltage is extracted by the symmetrical pair of resistors R.

At the DC bias point, the transistors T_a and T_b are equivalent (by symmetry) to a single transistor with gate and source connected. The transistors T_b are not saturated. Their output conductance G_{mdb} is proportional to the source transconductance G_{msa} of T_a:

$$G_{mdb} = PG_{msa} = nPG_m, \tag{5.71}$$

with

$$P \triangleq I_{specb}/I_{speca}, \tag{5.72}$$

where I_{spec} is the specific current of the transistors defined by (5.39). The output conductance of T_b creates a conductance $G_S = G_{mdb}/2$ in parallel with C_S. This conductance has been introduced in the expression (5.55) of Z_c to obtain the locus of $Z_c(G_m)$ plotted in Fig. 5.18b. This locus is no longer a circle, because Z_c is no longer a bilinear function of G_m. However, for P sufficiently small, it follows the circular locus of the basic circuit within an acceptable range of G_m.

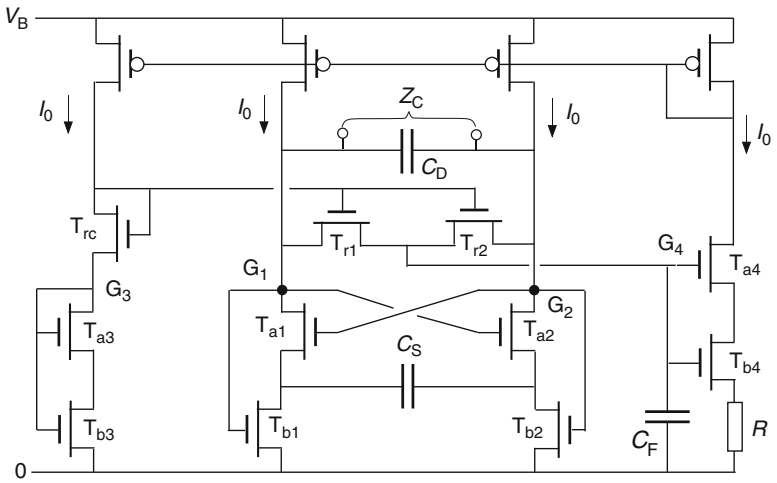

Fig. 5.19 Implementation of the parallel-resonance oscillator without resistors. An amplitude regulator is included [13]

In the calculation of this locus, the resistance R has been assumed to be sufficiently large to create negligible losses. However, the stability of the DC bias puts an upper limit to this value. Indeed, the gain of the amplifier formed of T_a degenerated by T_b and loaded by R must be smaller than unity. This condition can be expressed as

$$R < \frac{n(1+1/P)}{G_{mmmax}}, \tag{5.73}$$

where G_{mmax} is the maximum value of G_m (corresponding to a maximum of bias I_0) for which stability must be ensured. Thus, P must be very small to allow for a large value of R.

An alternative solution for biasing the circuit without resistors is illustrated in Fig. 5.19 [13].

The gate of T_b is now connected directly to the drain of the corresponding transistor T_a. But T_b is also active, which modifies the locus in a way that depends on the parameter P defined by (5.72), and by the mode of operation of the transistors. Figure 5.20 shows in full lines the loci of $Z_c(G_m)$ for T_a and T_b in strong inversion, and for T_a and T_b in weak inversion, for $P = 0.01$. The circular locus of the basic circuit is shown in thin line for comparison.

The maximum negative resistance is reduced, especially for weak inversion. Now, operating T_a in weak inversion is interesting to minimize the critical bias current, but there is no reason to operate T_b in weak inversion.

Intermediate loci for T_a in weak inversion and T_b in strong inversion are shown in dotted lines, further characterized by the parameter S defined by

$$S \triangleq \frac{I_{specb}}{\omega C_D n U_T}. \tag{5.74}$$

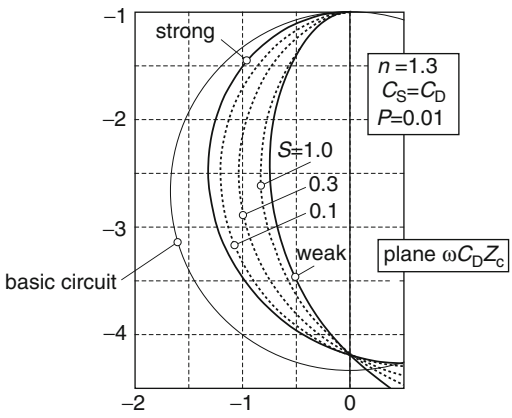

Fig. 5.20 Loci $Z_c(G_m)$ for the circuit of Fig. 5.19. *Full line*: transistors T_a and T_b all in weak inversion or strong inversion; comparison with the basic circuit in *thinner line*. *Dotted lines*: T_a in weak inversion, T_b in strong inversion with three values of parameter S

The circuit of Fig. 5.19 includes an amplitude regulator. For this purpose, two pairs of transistors matched to the main pairs T_a–T_b have been added. The same current I_0 flows through all four branches. The common mode voltage of the main pair is extracted by transistors T_{r1} and T_{r2} matched to their biasing transistor T_{rc}. The pair $T_{a3} - T_{b3}$ is identical to the main pairs. Hence, in absence of oscillation, all four gate nodes G_1–G_4 are at the same potential. But the specific current of both transistors of the fourth pair is *K-times larger* than that of the main pairs. This difference must be compensated by a voltage drop RI_0 across resistor R.

If all pairs are in weak inversion, the current needed to produce this voltage drop in absence of oscillation is

$$I_0 = I_{0start} = U_T \ln K/R. \tag{5.75}$$

If this current is larger than the critical current, the oscillation starts growing. The rectifying effect of transistors T_{r1} and T_{r2} lowers the DC voltage at node G_4, which reduces I_0, until equilibrium is reached. If these transistors are in weak inversion, the transfer characteristics of the regulator are

$$\frac{I_0}{I_{0start}} = 1 - \frac{\ln[\mathrm{I_{B0}}(|V_{in}|/2U_T)]}{n|\mathrm{n}K}, \tag{5.76}$$

where I_{B0} is the modified Bessel function of order zero. These characteristics are plotted in Fig. 5.21 for several values of the ratio K.

An example of the current-amplitude relationship for the oscillator in weak inversion (according to Fig. 5.15 where V_{in} was normalized to nU_T instead of U_T) is plotted in dotted line. Stable oscillation occurs at its intersection with the regulator curve.

The specific current of T_{r1} and T_{r2} must be much smaller than that of their biasing transistor T_{cr} to limit the equivalent loss conductance G_D due to these transistors.

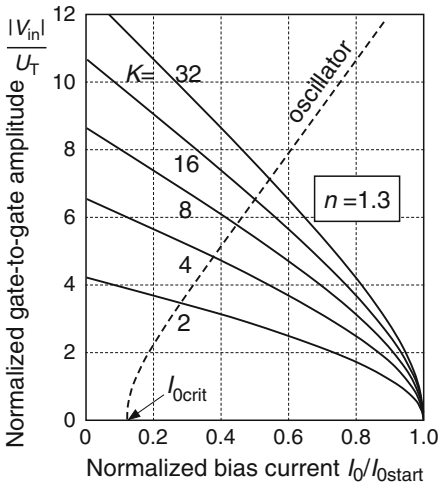

Fig. 5.21 Regulator characteristics for various values of ratio K. Stable oscillation occurs at the intersection with the oscillator characteristics (example in *dotted line*)

5.5 Series-Resonance Oscillator

5.5.1 Basic Circuit and Linear Analysis

A minimum of two active transistors is required to obtain a negative resistance that is essentially real. Among several possibilities, the symmetrical circuit illustrated in Fig. 5.22a was already known with vacuum tubes [14]. A bipolar version was used in first prototypes of electronic watches.

The transistors are assumed to be in the common substrate connected to the negative power supply rail zero. The corresponding negative-resistance circuit producing the impedance Z_c is shown in part (b) of the figure. Without capacitors C_P and C_L, or for DC signals, this impedance would be

$$Z_c|_{\omega=0} = \frac{2 - G_m R_L}{G_{ms}} = \frac{1}{n}\left(\frac{2}{G_m} - R_L\right), \tag{5.77}$$

where G_m and $G_{ms} = nG_m$ are the gate and source transconductances of each of the two transistors, and n is the slope factor. Thus, the role of the transistors is to transform the positive load resistance R_L into a negative resistance.

But the capacitance C_P includes the unavoidable parallel capacitance C_0 of the resonator defined by (5.2). To avoid oscillation of the circuit alone (without the motional impedance Z_m of the resonator), a compensating load capacitance C_L is necessary. It can be shown that the condition to avoid this parasitic oscillation at all values of $G_m > 0$ is

$$C_L > C_P/n. \tag{5.78}$$

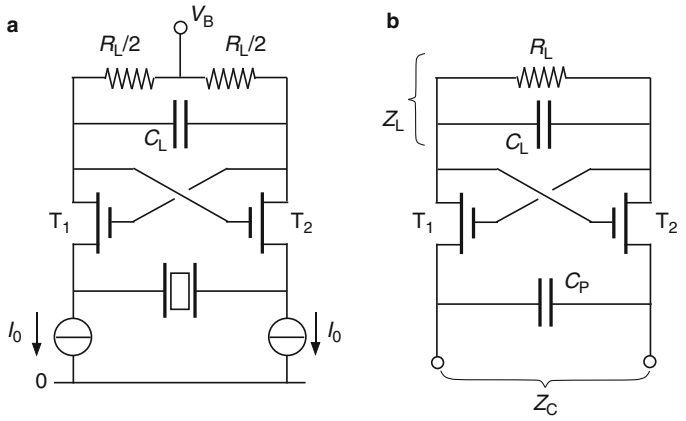

Fig. 5.22 Symmetrical oscillator for series resonance: (**a**) complete oscillator; (**b**) negative-resistance circuit without bias sources

The overall circuit impedance including possible losses associated with C_P can be expressed as

$$Z_c = \frac{Z_P(2 - G_m Z_L)}{2 + G_m(nZ_P - Z_L)}, \tag{5.79}$$

which is again a bilinear complex function of the transconductance G_m. For Z_P purely capacitive, the corresponding circular locus of $Z_c(G_m)$ is centered on the imaginary axis.

There is no reason to fulfill condition (5.78) with a margin larger than the maximum spread of capacitance values. Thus, the analysis will be simplified by assuming that this *condition is just fulfilled*, with $C_L = C_P/n$. The real and imaginary parts of Z_c can then be expressed as

$$\mathrm{Re}(Z_c) = \frac{1}{\omega C_P} \cdot \frac{AB(2B + 2/B - A)}{4 + (2B - A)^2} \tag{5.80}$$

$$\mathrm{Im}(Z_c) = \frac{1}{\omega C_P} \cdot \frac{4(AB - B^2 - 1)}{4 + (2B - A)^2}, \tag{5.81}$$

where A and B are the normalized transconductance and the normalized bandwidth, defined by

$$A \triangleq \frac{G_m}{\omega C_L} \quad \text{and} \quad B \triangleq \omega C_L R_L = \omega C_P R_L / n. \tag{5.82}$$

The normalized locus of $Z_c(G_m)$ is represented in Fig. 5.23 for several values of B.

It starts at $-j$ (thus $Z_c = 1/j\omega C_P$) for $G_m = 0$ and ends at $-B$ (thus $Z_c = -R_L/n$) for $G_m = \infty$. Notice that the imaginary part of the impedance is positive (inductive

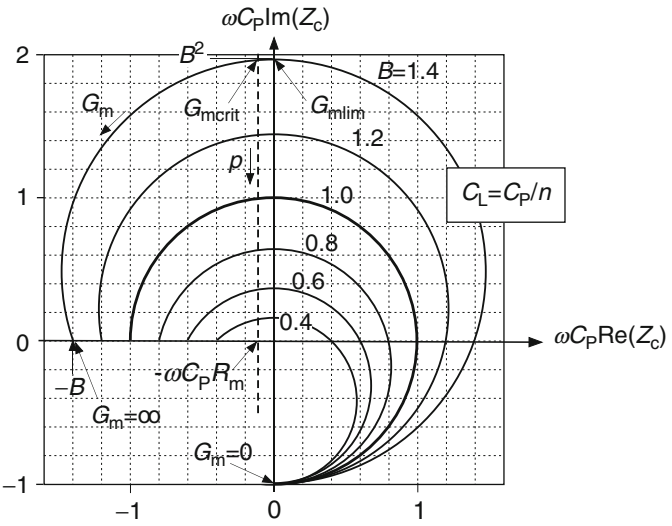

Fig. 5.23 Locus of $Z_c(G_m)$ for the series-resonance oscillator, for several values of the normalized bandwidth B. A possible locus of the motional impedance $-Z_m(p)$ is also represented in *dotted line*

impedance) when its real part is negative. The motional impedance must therefore be capacitive, which corresponds to a frequency of oscillation lower than the motional frequency ω_m ($p < 0$).

In order to obtain a negative resistance, the transconductance must reach a value G_{mlim} that is easily obtained from (5.80):

$$G_{mlim} = \frac{2\omega C_P}{n}(B + 1/B). \tag{5.83}$$

For a given value of the parallel capacitance C_P, this limit value is minimum for $B = 1$, that is, for $R_L = n/\omega C_P$. The imaginary value of Z_c at this limit obtained from (5.81) is

$$\mathrm{Im}(Z_c)|_{\mathrm{Re}(Z_c)=0} = \frac{B^2}{\omega C_P}. \tag{5.84}$$

If the motional resistance R_m is much smaller than the load resistance R_L (as depicted in the figure), the imaginary part of the impedance can be approximated by (5.84). Then, according to (5.13), the pulling is

$$p_c = -\frac{C_m}{2C_P}B^2. \tag{5.85}$$

When R_m approaches R_L, p_c tends to zero. This is only true if $C_L = C_P/n$, as is assumed in this analysis. If $C_L > C_P/n$, then p_c is increased and remains always negative.

The critical transconductance is larger than G_{mlim}. If $R_m \ll R_L$, it can be approximated by

$$G_{mcrit} = \frac{2}{R_L}\left(1 + B^2 + \frac{nR_m}{R_L}\right) = \frac{2\omega C_P}{n}\left(B + \frac{1}{B} + \frac{1}{2Q|p_c|}\right). \tag{5.86}$$

Thus, if $Q|p_c|$ is sufficiently large, then $G_{mcrit} \cong G_{mlim}$; the critical transconductance no longer depends on the quality factor.

5.5.2 *Amplitude of Oscillation*

As soon as the transconductance G_m exceeds the critical value given by (5.86), the amplitude of oscillation grows. The drain current of each transistor is then composed of the DC current I_0 imposed by the biasing source and an AC component I_1.

If the parallel capacitance C_P is not too large, most of the AC current I_1 flowing through the transistors is the motional current of the resonator; thus, this current remains approximately sinusoidal. Therefore, the gate voltage remains also sinusoidal, but the source voltage V_S is distorted. The effective transconductance is reduced, and more bias current I_0 is needed to further increase I_1, until it reaches its maximum value equal to I_0 (100% modulation of bias current I_0).

According to the EKV model [8, 9], the drain current in saturation depends on a combination of gate and source voltages, and on threshold voltage V_{T0}:

$$V_c = V_G - V_{T0} - nV_S \tag{5.87}$$

with a corresponding transconductance $G_{mc} \equiv G_m$. Assuming that the AC current I_1 is perfectly sinusoidal, the transconductance for the fundamental component is given by

$$G_{mc(1)} = \frac{|I_1|}{|V_{c(1)}|} = \frac{m_i I_0}{|V_{c(1)}|(m_i)}, \tag{5.88}$$

where $|V_{c(1)}|$ is the amplitude of the fundamental component of V_c produced by I_1 and $m_i = |I_1|/I_0$ is the index of current modulation. Stable oscillation is obtained for $G_{mc(1)} = G_{mcrit}$. Combining (5.88) with (5.36) yields

$$\frac{I_0}{I_{0critmin}} = \frac{1}{m_i} \cdot \frac{|V_{c(1)}|(m_i)}{nU_T}. \tag{5.89}$$

The control voltage V_c is a logarithmic function of the current in weak inversion, and a square root function in strong inversion. Its fundamental component is thus easily computed numerically. Results are plotted in Fig. 5.24. Strong inversion is again characterized by the inversion coefficient IC_0 at the critical current, as defined by (5.38).

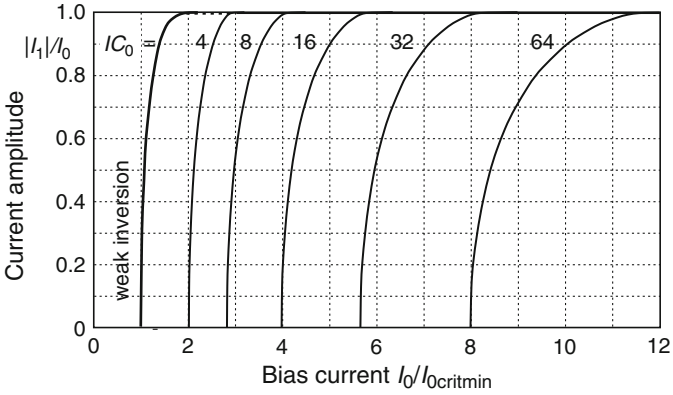

Fig. 5.24 Amplitude of oscillation of the series-resonance oscillator, assuming sinusoidal currents in the transistors

The AC current I_1 saturates to I_0 as soon as I_0 exceeds its critical value by about 50%. The assumption of a sinusoidal AC current would only be valid if the impedance of C_P were much larger than the motional impedance Z_m. Otherwise, harmonic current components created by the distortion of the source voltages can flow through C_P. As a result, I_1 tends to a square wave of amplitude I_0 for $I_0 \gg I_{0crit}$.

5.5.3 Phase Noise

The calculation of the excess phase noise for the linear case yields, for $4Qp_c \gg 1$,

$$\gamma_0 = (1+B^2)\frac{1+n\gamma_t}{n^2}\cdot\frac{R_L}{R_m}. \tag{5.90}$$

Since the load resistance R_L must be chosen sufficiently larger than the motional resistance R_m to have an acceptable margin of negative resistance, γ_0 is much larger than that for the Pierce and for the parallel-resonance oscillators discussed previously. Furthermore, the analytic approach based on the ISF for large amplitude is only feasible for $B \ll 1$, that is, for $G_{mcrit} \gg G_{mcritmin}$ according to (5.86). Therefore, phase noise will not be discussed any further for this type of oscillator.

5.5.4 Practical Implementation

Complementary voltage signals can be extracted at the drains by simple voltage followers. However, this circuit is by nature a current-mode oscillator, and the signal is best extracted by current mirrors in series with $R_L/2$, as illustrated in Fig. 5.25.

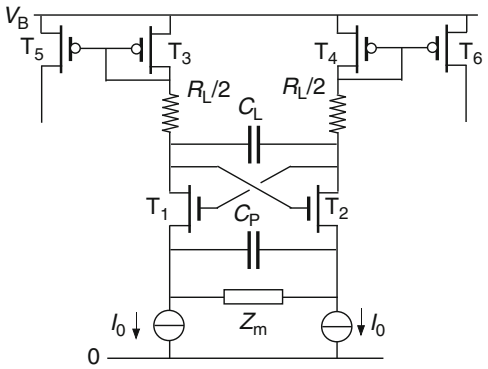

Fig. 5.25 Extraction of complementary oscillation currents

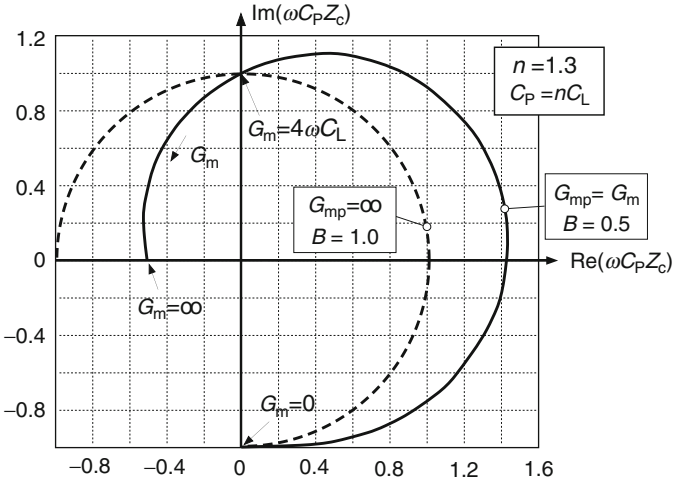

Fig. 5.26 Loci of $Z_c(G_m)$. In *dotted line*, basic circuit of Fig. 5.22a with $B = 1$; in *full line*, modified circuit of Fig. 5.25 with $B = 0.5$

Of course, the input transistors $T_3 - T_4$ of the mirrors modify the effective load resistance. An interesting situation occurs when the G_{mp} of these transistors is equal to the transconductance G_m of the active transistors, as is approximately the case if all transistors are in weak inversion. Indeed, according to (5.77), the DC negative resistance is then $-R_L$, independently of the value of G_m, which is compensated.

But this compensation does not take place for AC. Z_c is no longer a bilinear function of G_m; thus the locus is no longer a circle. It can be shown that the minimum transconductance producing a negative resistance is the same as that of the basic circuit, but that it occurs at $B = 0.5$ instead of $B = 1$. A comparison of the loci is illustrated in Fig. 5.26.

It shows that the imaginary part of Z_c (thus the pulling p_c for $R_m \ll R_L$) is the same in the two cases, but that its derivative is no longer zero for the circuit with mirrors. The pulling is thus more dependent on the value of R_m (thus on the quality factor Q).

5.6 Comparison and Conclusion

Let us identify the three types of oscillators discussed in this chapter as follows:

(a) The Pierce circuit of Sect. 5.3
(b) The parallel-resonance symmetrical circuit of Sect. 5.4
(c) The series-resonance symmetrical circuit of Sect. 5.5

With its single active device, (5.6) is the simplest of the three architectures. However, the symmetrical circuits (5.6) and (5.6) have the advantage of delivering symmetrical signals that are frequently needed in RF applications.

The critical currents for a given value of pulling p_c are given by (5.32) for (5.6) and by (5.60) for (5.6). Since (5.6) has two branches, the corresponding critical current is the same for the two circuits. But that of (5.6) can be reduced by a factor two by using the current-controlled complementary configuration of Fig. 5.10. As shown by (5.86) and (5.85), the critical transconductance for (5.6) has a minimum $4\omega C_P/n$ for $p_c = -C_m/2C_P$ and Q very large. It is more than for (5.6) and (5.6) for usual values of pulling p_c, but less sensitive to pulling. It may thus be smaller for very small values of p_c.

Pulling is always positive for (5.6) and (5.6) (frequency of oscillation above the mechanical resonant frequency of the resonator). It has a maximum value $C_m/2C_0$ limited by the minimum parallel capacitance C_0 given by (5.2). Pulling is always negative for (5.6) (frequency of oscillation below the mechanical resonant frequency) and has no fixed limit, except for an increase of current.

Since it exploits the high-impedance parallel resonance (or antiresonance), (5.6) is best suited for minimizing power consumption when no requirement is put on frequency pulling (that is, on the dependency of the frequency of oscillation on electrical components).

The maximum obtainable negative resistance is limited by the parameters of the resonator for (5.6). For (5.6), it is limited by the need to respect the stability condition (5.54) in spite of the capacitance mismatch. The maximum negative resistance of (5.6) is given by the load resistance R_L. It can in principle be increased to very high values, at the cost of an increase of pulling, current I_0, and of course supply voltage $V_B > R_L I_0/2$.

The symmetrical circuit (5.6) is better than (5.6) for phase noise if the phase shift $\Delta\phi$ between drain current and drain voltage is not negligible. Both are better than the series-resonance circuit (5.6).

Overall, although (5.6) is in principle best adapted to the series-resonance motional circuit of the resonator, its behavior is drastically degraded by the unavoidable parallel capacitor $C_P \geq C_0$. As an advantage, it delivers complementary currents with 100% modulation. These currents are approximately sinusoidal if $B = \omega C_L R_L \ll 1$.

References

1. G.W. Pierce, Proc. Am. Acad. Arts Sci. **59**, 81 (1923)
2. E. Vittoz, *Low-Power Crystal and MEMS Oscillators* (Springer, New York, 2010)
3. E. Vittoz, in *Proceedings of the 10th International Congress of Chronometry*, Geneva, 1979, pp. 131–140
4. E.A. Vittoz, M.G.R. Degrauwe, S. Bitz, IEEE J. Solid State Circ. **23**(3), 774 (1988)
5. J. Philips, K. Kundert, in *IEEE 2000 Custom Integrated Conference*, 2000, pp. 431–431
6. A. Hajimiri, T.H. Lee, IEEE J. Solid State Circ. **33**(2), 179 (1998). Corrections at page 928 of the same issue
7. J.K. Clapp, IRE Proc. **36**(3), 356 (1948)
8. C.C. Enz, F. Krummenacher, E.A. Vittoz, special issue of the Analog Integr. Circ. Signal Process. J. Low-Voltage Low-Power Des. **8**, 83 (1995)
9. C.C. Enz, E.A. Vittoz, *Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design*, 1st edn. (Wiley, New York, 2006)
10. E. Vittoz, J. Fellrath, IEEE J. Solid State Circ. **12**(3), 224 (1977)
11. E.A. Vittoz, in *Sub-Threshold Design for Ultra Low-Power Systems*, ed. by A. Wang, B.H. Calhoun, A.P. Chandrakasan, 1st edn., Series on Integrated Circuits and Systems (Springer, New York, 2006), pp. 49–74
12. D. Aebischer, et al., IEEE J. Solid State Circ. **32**(7), 999 (1997)
13. D. Ruffieux, in *Proceedings of the European Solid-State Circuits Conference (ESSCIRC '02)*, Firenze, 2002, pp. 85–88
14. H.J. Reich, *Functional Circuits and Oscillators* (D. Van Nostrand Co, Princeton, 1961). pp. 80–85

Chapter 6
5.4 GHz, 0.35 µm BiCMOS FBAR-Based Single-Ended and Balanced Oscillators in Above-IC Technology

Éric Tournier

Abstract For the last few years, one of the main challenges in circuit design has been the integration of frequency references in applications where phase noise requirements are very stringent. To overcome the usual limitation of (Bi)CMOS integrated circuits (ICs) phase noise, mainly due to the low *Q*-factor of standard integrated passive devices (*R*, *L*, *C*) inherent to low resistivity substrate, a solution has been to use BAW resonators. Indeed, thin film BAW resonators based on piezoelectric material, generally AlN or ZnO, sandwiched between two metallic electrodes, exhibit a high *Q*-factor, can handle high power, and can operate at high frequencies (above 10 GHz), while keeping a small size and being compatible with (Bi)CMOS IC processes. The very first oscillators using FBAR and SMR resonators were designed with separately wire-bonded resonators connected to the IC circuit. This chapter deals with the world premiere realization of two 5-GHz FBAR-based oscillators, where the FBAR is directly integrated above the IC with some further process steps, compatible with (Bi)CMOS. A single-ended and a balanced version were designed. The circuits were implemented in a 0.35-µm SiGe BiCMOS process from AMI Semiconductor. From the obtained results, we show that post-processing the FBAR directly over the IC eliminates much of the parasitics and modelling issues associated with bondwires. Furthermore, it reduces the circuit area. The single-ended and balanced oscillators are based on the Colpitts configuration and achieve respectively a state-of-the-art phase noise performance (at the time of design) of −117.7 dBc/Hz and −121 dBc/Hz at 100 kHz offset from the 5.4-GHz carrier frequency. The balanced version allows direct driving of balanced dividers and mixers without the need of a single-ended to balanced converter. Some comparisons are also made with standard *LC* balanced oscillators.

É. Tournier (✉)
LAAS/CNRS, Université de Toulouse, France
e-mail: eric.tournier@laas.fr

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_6,

6.1 Introduction

Oscillators are a key function of any communication systems. They generate sinusoidal signals at frequencies where the information spectrum needs to be translated, for example, to fit into the bandwidth of a given channel, or to switch between different users of a frequency division multiplexing (FDM) protocol.

In case of commercial applications, low phase noise, high output power, high integration, low cost and low consumption are often required. While (Bi)CMOS technologies are the main candidates to integrate most of the hardware, these technologies suffer from low quality passive devices, especially inductors, which do not allow the design of resonator with a sufficient quality factor. Then, the choice is either to use these devices to the detriment of the performance or to use external stand-alone resonators to the detriment of integration level. Even the quite popular SAW device, widely used in the current mobile communication systems, is likely to become inadequate for high-frequency applications above 5 GHz due to degraded performance and poor electrical power-handling capabilities [1]. Moreover, technology steps of SAW technology are incompatible with standard (Bi)CMOS technologies.

In recent years, among various approaches and breakthroughs in integration processes to allow more and more miniaturization of RF systems, the introduction of bulk acoustic wave (BAW) devices has been decisive. BAW resonators, with their excellent behaviour at high frequency, their high quality factor and their compatibility with (Bi)CMOS technologies, are the main candidate to reach all the design requirements, together with a high integration level.

At the time of design, some realizations of oscillators based on BAW devices had been reported, but all were done with discrete devices wire-bonded on the main circuit afterwards [2–4]. This chapter describes the design of two completely integrated FBAR-based oscillators, featuring for the first time a 5-GHz operating frequency, in their single-ended [5] and balanced [6] topologies. For the purpose of integration, an above-IC technique was used, which allowed processing BAW technology steps as a natural extent of (Bi)CMOS ones [7], leading to high-frequency, high-performance monolithic oscillators.

Our initial goal was to demonstrate the benefit of having above-IC BAW instead of *LC* resonators. With this goal in mind, we did not use very sophisticated design techniques, to let all the improvement benefit on the BAW resonators and their above-IC connections. The work described in this chapter has been done in the frame of the European project MARTINA, during the years 2003–2006.

After a few words on oscillators in transceivers, and some others on *LC* and BAW resonators, we present the above-IC technology used for the oscillators' integration. Then, we introduce the single-ended and balanced topologies, with some informations on the design steps, and we give comments on the obtained performances of the designed oscillators. Finally, we compare them with some more classical balanced oscillators based on *LC* resonators.

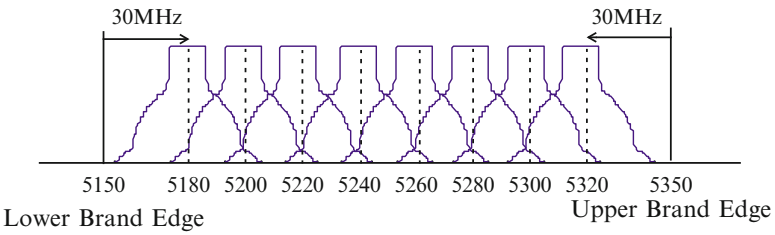

Fig. 6.1 Lower and middle U-NII bands: 8 carriers in 200 MHz/20 MHz

As a targeted standard to design such oscillators, we based our work on the lower band (200 MHz) of the 802.11a standard [8], an overview of which is given in Fig. 6.1.

6.2 Resonators and Oscillators in Transceivers

6.2.1 Resonators' Use in Transceivers

Figure 6.2 gives an example of a typical direct conversion transceiver. I and Q are in-phase and quadrature signals from/to a digital signal processor (DSP), which controls the constellation of the modulation through mixing with a local oscillator (LO) and its 0 degree and 90 degree outputs. The functions (shown in bold) where BAW resonators can be used are filters, duplexer, and local oscillators.

In this chapter, we are focusing on the local oscillator, and Fig. 6.3 gives an example of using a BAW resonator as a reference for such an oscillator in a phase-locked loop (PLL). Indeed, as the BAW resonator is very selective, it will not allow any frequency shift in the oscillator output. A *LC* oscillator is locked to the BAW oscillator through a wide-bandwidth PLL allowing to set the output frequency to arbitrary frequencies.

6.2.2 Oscillators' Use in Transceivers

In a transceiver, local oscillators generate sinusoids that are used to translate the spectrum of the signal of interest to a different frequency. Most of the time, a mixer is employed to implement this translation also referred to as heterodyning.

If $x(t)$ is the signal of interest, $X(f)$ its spectrum, and f_o the frequency generated by a local oscillator $o(t)$ that, in the best case, features a time-varying voltage

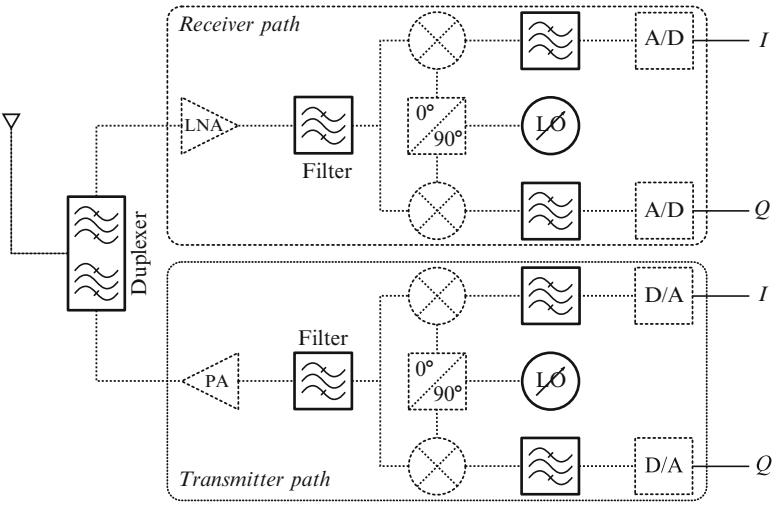

Fig. 6.2 A typical direct conversion transceiver

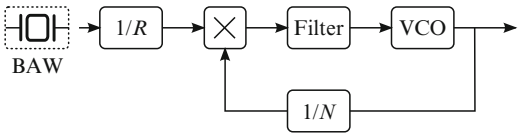

Fig. 6.3 A phase-locked loop on a BAW stabilized reference

given by $V_o \cos(2\pi f_o t)$, the mixing of $x(t)$ with $o(t)$ leads to a double-sideband spectrum:

$$\begin{aligned}\mathscr{F}(x(t)\,o(t)) &= X(f) \star \left\{ \frac{V_o}{2} \left[\delta(f-f_o) + \delta(f+f_o)\right] \right\} \\ &= \frac{V_o}{2} \left[X(f-f_o) + X(f+f_o)\right] \end{aligned} \tag{6.1}$$

The baseband spectrum $X(f)$ is simply translated around f_o. Obviously, if the local oscillator is not stable and exhibits amplitude and phase noises ($v(t)$ and $\varphi(t)$, respectively), the spectrum after translation will be degraded. In this case, $o(t)$ can be written as : $(V_o + v(t)) \cos(2\pi f_o t + \varphi(t))$. As the amplitude noise can be quite easily cancelled by a proper limiter or amplitude control circuit, the phase noise remains the only degradation of the local oscillator: $o(t) = V_o \cos(2\pi f_o t + \varphi(t))$. By inspecting its power spectrum density:

$$S_o(f) = \frac{V_o^2}{4} \left[\delta(f-f_o) + S_\varphi(f-f_o) + \delta(f+f_o) + S_\varphi(f+f_o)\right] \tag{6.2}$$

we can now see the influence of the spectral density S_φ of phase variation which directly impacts the transceivers' quality, regarding signal-to-noise ratio for analogue transmission, or bit error rate for "digital transmission."[1]

6.2.2.1 Figure of Merit for Phase Noise Design

The more significant oscillator characteristics are frequency of oscillation, frequency tuning range, frequency stability (phase noise and jitter), purity (harmonics), amplitude stability, output power and power dissipation. In order to ease the comparison between different oscillators, an appropriate figure of merit (FoM) is needed. It must include most of the characteristics of the oscillator, keeping in mind that a trade-off between some of them is usually needed in order to optimize a given parameter at the expense of another one. This FoM is usually given as [9]:

$$\mathrm{FoM} = 20\log\left(\frac{f_o}{f_{\mathrm{offset}}}\right) - 10\log\left(\frac{P_{\mathrm{diss}}}{1\,\mathrm{mW}}\right) - L(f_{\mathrm{offset}}) \tag{6.3}$$

An appropriate FoM should remain unchanged if a simultaneous modification of two dependant parameters occurs in the proportion of their dependance.

The definition (6.3) takes into account the phase noise at a given frequency offset f_{offset} from the carrier f_0, which is the main contribution of the FoM: the lower the phase noise, the higher the FoM. Its value is then balanced with the $20\log(f_o/f_{\mathrm{offset}})$ term which ensures that watching the phase noise at a different offset will not change the FoM as long as the phase noise exhibits a classical -20 dB/dec slope. The $P_{\mathrm{diss}}/1$ mW term normalizes the FoM to a power dissipation of 1 mW.

Some properties such as the oscillator output power are, however, not included in this widely used FoM.

6.2.2.2 General Oscillator Design

To design an oscillator, we need a resonator featuring a resonant frequency equal to the expected oscillating frequency, together with an amplifier to compensate for the losses of the resonator.

The well-known Barkhausen criteria are used to ensure that oscillations will start and remain stable. These conditions can be expressed under many forms: with transfer functions, with impedances, with *S*-parameters and so on, all being in fact equivalent.

If we use here, for example, the transfer function formalism in a feedback approach, well adapted for transmission type oscillations, the conditions can be expressed as (Fig. 6.4a):

$$|A(\omega_0)\,Q(\omega_0)| = 1 \tag{6.4}$$

$$\angle A(\omega_0) + \angle Q(\omega_0) = 0 \tag{6.5}$$

[1] More exactly "analog transmission of digital information."

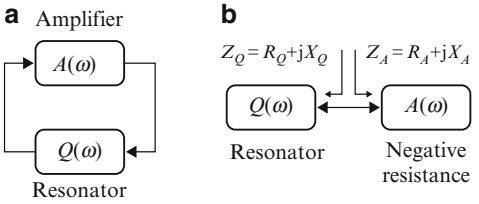

Fig. 6.4 Oscillators in a feedback approach (**a**) and in a negative resistance approach (**b**)

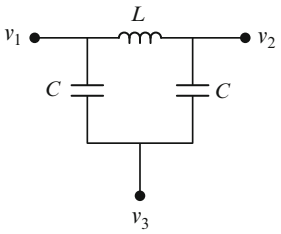

Fig. 6.5 Colpitts/Pierce/Clapp *LC*-resonator

where $A(\omega)$ is the amplifier transfer function and $Q(\omega)$ is the resonator transfer function and ω_0 its resonant frequency.

In a negative resistance approach, more adapted to oscillations in reflexion mode, and speaking in Z-parameters terms, we have (Fig. 6.4b):

$$R_A(\omega_0) = -R_Q(\omega_0) \tag{6.6}$$

$$X_A(\omega_0) = -X_Q(\omega_0) \tag{6.7}$$

where $R_A(\omega)$ (resp. $R_Q(\omega)$) is the resistive part and $X_A(\omega)$ (resp. $X_Q(\omega)$) the reactive part of the negative resistance (resp. resonator) impedance. Note that at the very beginning of the oscillation, we will need to have either $|A(\omega_0)\,Q(\omega_0)| > 1$ or $R_A(\omega_0) > -R_Q(\omega_0)$ (and not only equality) for the oscillations to take place and their amplitude to grow up to their final value where the equalities (6.6) and (6.7) will be valid.

6.2.3 LC Resonators

In their simplest form, resonators can be made with inductors (L) and capacitors (C). *LC* resonators can be designed in many ways, the most common ones having a name: Colpitts, Pierce, Clapp, Hartley, Armstrong… Well, with time and according to authors, all the (L,C) topologies behind these names change, and it is difficult to find a unified definition. By inspecting a quite old reference [10], we found a possible one. Figure 6.5 shows the resonator used in Colpitts/Pierce/Clapp, with one inductor and two capacitors. The signal goes from v_1 to v_2 while v_3 is tied to ground for the Pierce, from v_3 to v_1 while v_2 is tied to ground for the Colpitts, from v_1 to v_3 while v_2 is tied to ground for the Clapp (v_1 and v_2 are interchangeable).

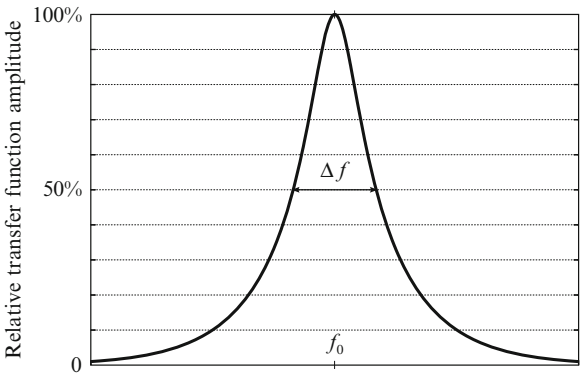

Fig. 6.6 *Q*-factor illustration

To ensure that the oscillations will be as close as possible to a perfect sinusoid, the selectivity of the resonator needs to be the highest possible, so that the oscillation condition will be met only for a tiny bandwidth. The selectivity is directly correlated to the *Q*-factor of the resonator. But the main problem with integrated *LC* resonators is that their *Q*-factor is very low, leading to poor phase noise performances. Let us remind that the *Q*-factor is defined from a physical point of view as:

$$Q = 2\pi \times \frac{\text{Energy stored}}{\text{Energy dissipated per cycle}} \tag{6.8}$$

or for a more usable definition:

$$Q = \frac{f_0}{\Delta f} \tag{6.9}$$

where Δf is, for example, the transfer function bandwidth at 50% as shown in Fig. 6.6.

6.2.4 Phase Noise Versus Q-Factor

By following the simple theory of Leeson [11], we can see that the oscillator phase noise is directly related to the resonator *Q*-factor by the expression:

$$F(f_m) = \frac{f_0}{2\sqrt{2}Q f_m} + \text{residual phase noise of the negative resistance} \tag{6.10}$$

f_m is the frequency offset from the carrier, where the phase noise power is measured in a 1-Hz bandwidth. This simple equation shows that better phase noise will be obtained with higher resonator *Q*-factor.

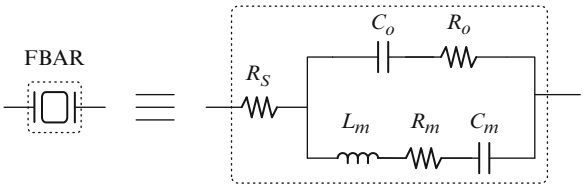

Fig. 6.7 BAW resonator modified BVD model

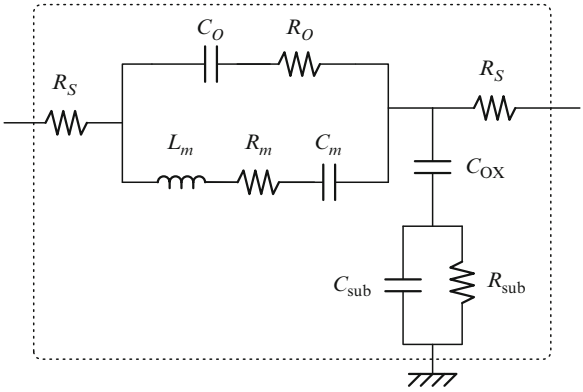

Fig. 6.8 BAW resonator modified BVD model with substrate effect

6.2.5 BAW Resonators

BAW resonators can be modelled with an *RLC* equivalent circuit. The BVD (Butterworth–Van Dyke) model, in its original [12] and modified version [13], is the most widely used in design because of its simplicity (Fig. 6.7).

R_s models the ohmic losses in the electrodes. C_o is the main capacitance defined by the electrodes and the dielectric. R_o represents dielectric losses (a few tenth ohm), and exists only in the modified version of the BVD model. R_m, L_m and C_m are respectively the acoustic resistance, inductance and capacitance. This model is valid only in the vicinity of the main resonant mode and only if the width of the resonator is much higher than its thickness.

A version that includes substrate losses can also be defined as in Fig. 6.8.

The typical impedance module of this model versus frequency is given in Fig. 6.9, where f_s and f_p are respectively the series and the parallel resonance, given after a few assumptions by [10]:

$$f_s \approx \frac{1}{2\pi\sqrt{L_m C_m}}$$

$$f_p \approx \frac{1}{2\pi\sqrt{L_m\left(\dfrac{C_o C_m}{C_o + C_m}\right)}} \approx f_s\left(1 + \frac{C_m}{2C_o}\right)$$

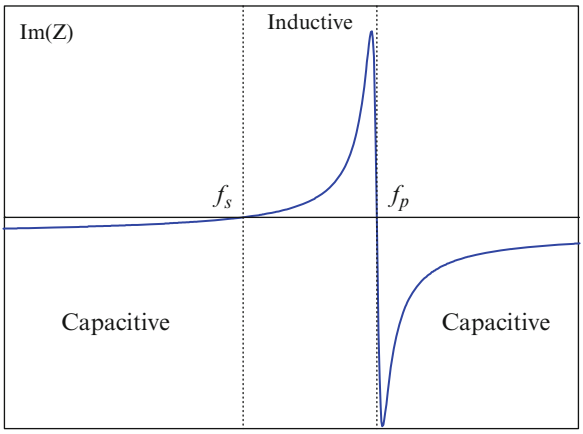

Fig. 6.9 Series and parallel resonance frequencies

From the graph on Fig. 6.9, we can see that the resonator acts as a capacitor for most of the frequency range, except in the range $[f_s; f_p]$ where it acts as an inductor. Then, we can replace the inductor of the resonator topology given in Fig. 6.5 by a BAW resonator working in its $[f_s; f_p]$ range.

With the thin electrode approximation, we can determine the values of each element of the BVD model as follows, with only the mechanical and geometrical characteristics [14]:

$$C_o = \frac{\varepsilon_r \, \varepsilon_0 \, A}{d} \tag{6.11}$$

$$R_m = \frac{\pi \, \eta \, \varepsilon_r \, \varepsilon_0}{8 \, k_t^2 \, \rho \, A \, \omega \, V_a} \tag{6.12}$$

$$L_m = \frac{\pi^3 \, V_a}{8 \, \varepsilon_r \, \varepsilon_0 \, \omega_r^3 \, k_t^2 \, A} \tag{6.13}$$

$$C_m = \frac{8}{\pi^2} \, k_t^2 \, C_o = \frac{8 \, \varepsilon_r \, \varepsilon_0 \, \omega_r \, A}{\pi^3 \, V_a} \, k_t^2 \tag{6.14}$$

where:

A	is the resonator area
ε_r	relative permittivity of piezoelectric material
ε_0	vacuum permittivity
ρ	piezoelectric material density
η	viscosity
ω	pulsation
k_t^2	coupling coefficient

R_s and R_o are losses of a few tenth ohm.

The coupling coefficient is defined as:

$$k_t^2 = \frac{\frac{\pi}{2}\frac{f_s}{f_p}}{\tan\left(\frac{\pi}{2}\frac{f_s}{f_p}\right)} \approx \frac{\pi^2}{4} \cdot \frac{f_p - f_s}{f_p} \tag{6.15}$$

This coupling coefficient is very important because it is directly proportional to the difference of resonant frequencies $f_p - f_s$. A large coupling coefficient will give the possibility to use the resonator on a larger range of frequencies, where its behaviour is inductive.

By measuring f_s and f_p, one will not obtain exactly the intrinsic coupling coefficient seen from the material, but the effective coupling coefficient seen from the electrodes:

$$k_{\text{eff}}^2 = \frac{\pi^2}{4} \cdot \frac{f_p - f_s}{f_p}$$

The coupling coefficient of harmonic resonances is given by:

$$k_{t,n}^2 = \frac{k_t^2}{n^2}$$

where n is the harmonic index. Because of the inverse square dependance to n, the usability of the resonator around its harmonic frequencies diminishes as the harmonic index increases: it will be, for example, very difficult to control the frequency of an oscillator based on a high harmonic index (this can also be seen as an advantage).

The Q-factor, which is very important as explained earlier in Sect. 6.2.4, can be expressed as:

$$Q_p = Q_s = \frac{L_m\,\omega}{R_m} = \frac{V_a^2\,\rho}{\omega\,\eta}$$

when ohmic losses are neglected. By taking into account the ohmic losses, the series Q-factor can be rewritten as:

$$Q_s = \frac{L_m\,\omega}{R_m + R_s}$$

In order to easily compare the performances of different resonators, a figure of merit derived from the Q-factor and the effective coupling coefficient is [15]:

$$\text{FoM} = \frac{Q}{2C_r} = k_{\text{eff}}^2\,Q$$

where:

$$C_r = \frac{C_o}{C_m}$$

Q-factors of BAW resonators, are consequently tens of times those of *LC* resonators.

6.3 Above-IC Technology

The basic configuration of a BAW resonator is a piezoelectric thin film sandwiched between two metal electrodes. At the mechanical resonance, the half wavelength of the acoustic wave corresponds to the total thickness of the stack. Actual resonators need to be mechanically isolated from the substrate on which they are built, in order to confine energy in the resonator itself and hence achieve high quality factors [16]. Figure 6.10 shows the two main configurations for realizing this isolation. The first, also known as film bulk acoustic resonator (FBAR), is built on a membrane suspended in air by its edges and manufactured over a sacrificial layer. The second configuration is the solidly mounted resonator (SMR), in which the acoustic impedance of the substrate is transformed to a very low value by a reflector made of a set of quarter wavelength sections of materials having large impedance ratios [7].

In this work, FBARs fabricated by surface micro-machining were used since they seem more appropriate for an above-IC integration. Indeed, since only a very thin isolating air gap is created underneath each resonator, the latter is placed very close to the IC. Low loss interconnections are hence much easier to achieve than if vias were to be implemented through a thick acoustic reflector. The heart of a BAW resonator is the piezoelectric thin film. Aluminium nitride (AlN) is

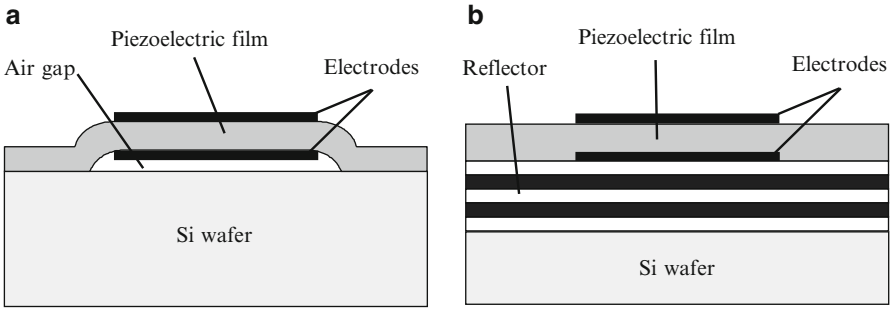

Fig. 6.10 FBAR (**a**) versus SMR (**b**) resonators. ©2006 IEEE. Reprinted, with permission, from [7]

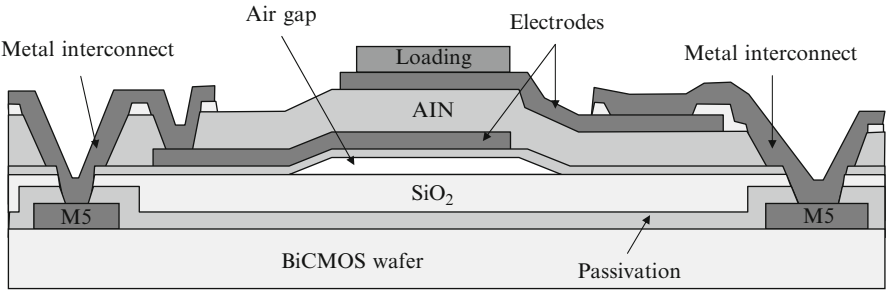

Fig. 6.11 Overview of the above-IC process. ©2006 IEEE. Reprinted, with permission, from [7]

most often preferred over competitor materials, such as ZnO or PZT, because it offers an excellent compromise between performance and manufacturability. Its coupling might not be very high compared to the other materials, but it is chemically very stable and has an excellent thermal conductivity and a low temperature coefficient [7].

FBAR resonators are post-processed over 0.35-µm BiCMOS SiGe wafers, from AMIS technology. First, silicon oxide is deposited over the passivated wafers and planarized by chemical–mechanical planarization in order to get a smooth and flat surface for the FBAR fabrication. A polymer sacrificial layer is deposited and patterned for defining each resonator's position, and a dielectric encapsulation layer is applied to protect the polymer. Next, the active part of the devices is built up, with the subsequent deposition and patterning of a Pt bottom electrode, the piezoelectric AlN layer and an Al top electrode. Via holes are then etched through the different dielectric layers down to the last metal level of the integrated circuit (M5), and a thick Al film is deposited and patterned to create interconnections between the BAW filter and the IC. Finally, the sacrificial layer is etched off for releasing the membranes. Figure 6.11 shows a schematic cross section of a FBAR connected to the last metal level of an IC wafer [7].

Figures 6.12 and 6.13 show the microphotograph of two FBAR resonators, a single-port one (grounded) and a two-port one (floating), taken at the time of the measurements of their *S*-parameters from which the BVD model has been deduced.

Only two-port resonators have been investigated in this work. Table 6.1 gives measurement values obtained from tens of resonators, and Fig. 6.14 shows the fundamental resonance together with order two and three resonances for a typical resonator.

These measurements led us to a BVD model (Fig. 6.7 and Table 6.2), and the extracted parameters have been used for designing forthcoming oscillators later in this chapter.

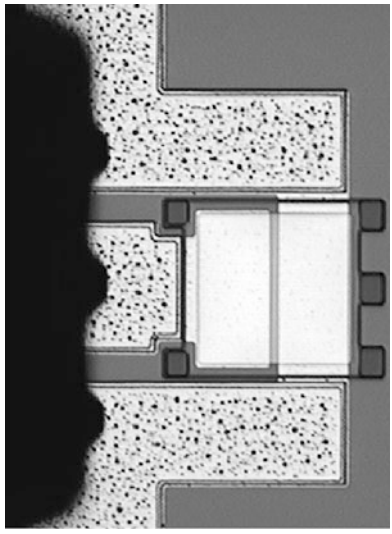

Fig. 6.12 One-port FBAR resonator (reflection type, $210\times285\,\mu m^2$)

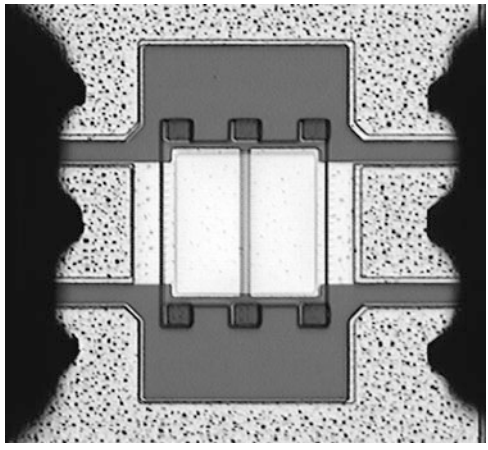

Fig. 6.13 Two-ports FBAR resonator (transmission type, $280\times260\,\mu m^2$)

Table 6.1 Measurements of two-port FBAR

	f_s (GHz)	f_p (GHz)	Coupling coefficient (%)	Q_s	Q_p
Mean	5.542	5.700	6.67	294	341
Standard deviation	0.1455	0.1545	0.22	55	115
Minimum	5.326	5.465	6.12	185	113
Maximum	5.863	6.041	7.09	396	553

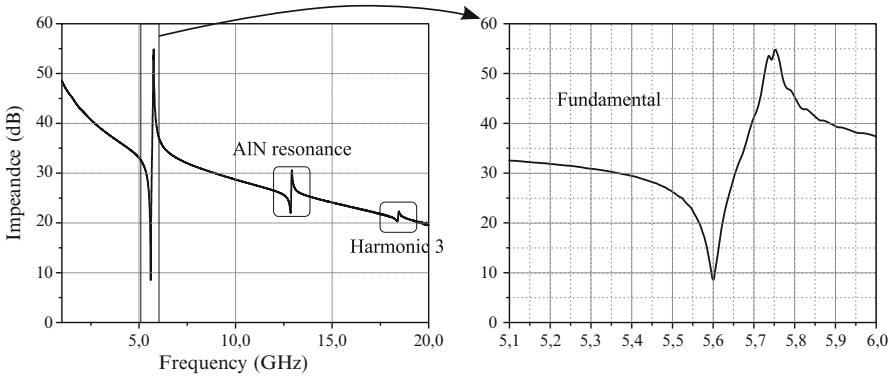

Fig. 6.14 The first three resonances of the FBAR

Table 6.2 BVD model of the FBAR resonator used

$R_m(\Omega)$	L_m(nH)	C_m(fF)	C_o(pF)	$R_o(\Omega)$	$R_s(\Omega)$
1.2	34.64	25.55	0.52	0.55	2.15

6.4 FBAR-Based Single-Ended and Balanced Oscillators

When working at very high frequency, there is a significant benefit in designing balanced functions: because of the inherent symmetry of both the topology and signals, design is most often based on relative values in case of balanced functions, instead of absolute values as in the case of single-ended ones; and we know how good the matching of identical components on a wafer can be, whereas a single component will have a high dispersion of its characteristics from its typical value through multiple runs of the same technology.

In this section, we present a single-ended [5] and a balanced version [6] of FBAR-based oscillators.

6.4.1 Single-Ended Version

The schematic of the single-ended oscillator is given in Fig. 6.15. Figure 6.16 shows the equivalent model of the oscillator where the FBAR is replaced by its BVD model around the resonant frequency.

As said before, in order to study the oscillator around its periodic steady state, we can use the model given in Fig. 6.17, where the equivalent negative resistance R_A of the transistor input impedance compensates for the equivalent resistance R_Q of the resonator, and the capacitive reactive part of the transistor input impedance $X_A = C_A$ compensates for the equivalent inductive reactive part $X_Q = L_Q$ of the resonator.

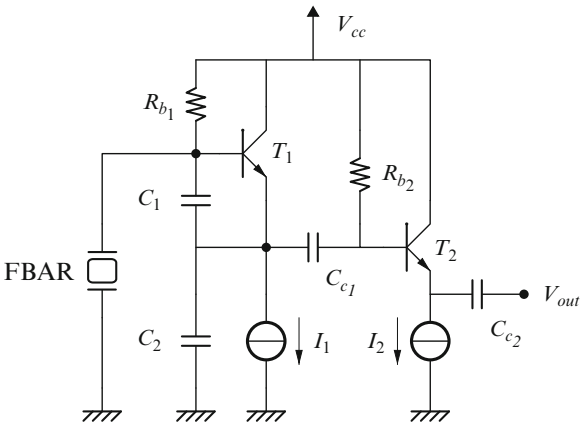

Fig. 6.15 Schematic of the single-ended oscillator

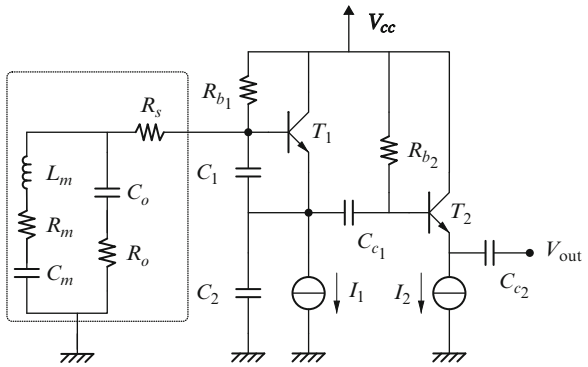

Fig. 6.16 Single-ended oscillator with the BAW BVD model

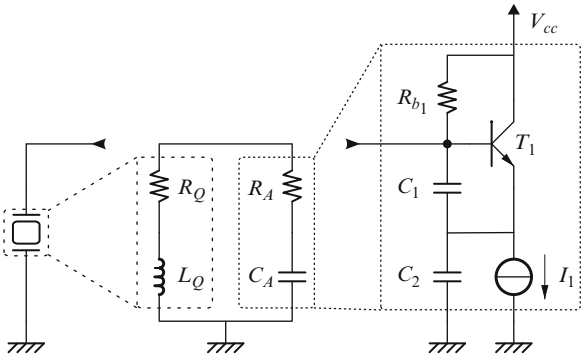

Fig. 6.17 Equivalent model at oscillation

By neglecting most of the parasitic components (low base-collector capacitance (with no Miller effect here), high collector resistance, high biasing resistance $R_{b_1}, \ldots$) and a base-emitter resistance $R_\pi \gg 1/(\mathrm{j}(C_1 + C_\pi)\,\omega)$ near the resonance, we can make a very simple model of the transistor topology (Fig. 6.18).

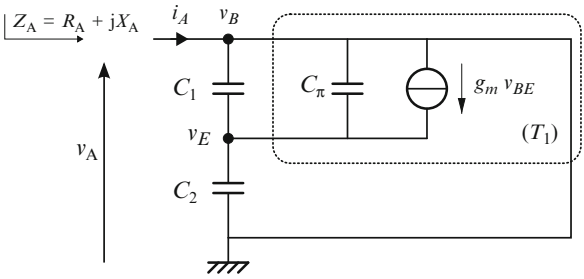

Fig. 6.18 Small signal input impedance of the transistor in the vicinity of the resonant frequency

We can write:

$$v_{BE} = \frac{i_A}{(C_1 + C_\pi)\,\mathrm{j}\,\omega} \tag{6.16}$$

$$v_A = \frac{i_A + g_m\, v_{BE}}{C_2\,\mathrm{j}\,\omega} + v_{BE} \tag{6.17}$$

where g_m is T_1 transconductance and C_π its base-emitter capacitance. Then replacing v_{BE} in (6.17) by its value (6.16) gives:

$$\begin{aligned}\frac{v_A}{i_A} = Z_A &= \frac{1}{C_2\,\mathrm{j}\,\omega} + \left[\frac{g_m}{C_2\,\mathrm{j}\,\omega} + 1\right]\frac{1}{(C_1 + C_\pi)\,\mathrm{j}\,\omega} \\ &= \frac{(C_1 + C_\pi + C_2)\,\mathrm{j}\,\omega}{C_2\,\mathrm{j}\,\omega\,(C_1 + C_\pi)\,\mathrm{j}\,\omega} + \frac{g_m}{C_2\,\mathrm{j}\,\omega\,(C_1 + C_\pi)\,\mathrm{j}\,\omega}\end{aligned}$$

We easily identify:

$$Z_A = \frac{-g_m}{C_2\,(C_1 + C_\pi)\,\omega^2} - \mathrm{j}\,\frac{C_1 + C_\pi + C_2}{C_2\,(C_1 + C_\pi)\,\omega} = R_A + \mathrm{j}\,X_A \tag{6.18}$$

The real part of the impedance seen at the input of the transistor T_1 can be considered as a negative resistance whose value is:

$$R_A = \frac{-g_m}{C_2\,(C_1 + C_\pi)\,\omega^2} \tag{6.19}$$

The reactive part can also be identified to an equivalent capacitor:

$$X_A = \frac{1}{C_A\,\mathrm{j}\,\omega} = \frac{1}{\dfrac{C_2\,(C_1 + C_\pi)}{C_1 + C_\pi + C_2}\,\mathrm{j}\,\omega} \Leftrightarrow C_A = \frac{C_2(C_1 + C_\pi)}{(C_1 + C_\pi + C_2)} \tag{6.20}$$

The negative resistance's absolute value can be increased at a given frequency with the decrease of C_1 or C_2, or the increase of g_m with the bias current I_1. As the C_1 and C_2 capacitors are generally fixed by the oscillation frequency, I_1 remains the only way to modify the negative resistance.

For oscillation start-up, the negative resistance must not just compensate for the resistive behaviour of the FBAR resonator, as it is the case to maintain oscillations when they have started: its absolute value must be higher.

The transistor T_2 acts as a buffer to isolate the resonator from the 50 Ω impedance of the measurement instruments. It is coupled to the oscillator through a small coupling capacitor C_{c_1}.

By studying Fig. 6.7 to which an equivalent load capacitor $C_L = C_A$ is connected and the resistor R_o is neglected, and after a few assumptions [10], we can write the equivalent resistance (real part of the impedance) of the resonator:

$$R_Q = R_s + R_m \left(1 + \frac{C_o}{C_L}\right)^2 \tag{6.21}$$

The absolute value of the negative resistance must be larger than this equivalent resistance in order to allow start-up of the oscillation. A larger amplitude, and a better phase noise, can be obtained by using a small capacitance C_2. However, a small C_2 will increase the equivalent resistance R_Q of the resonator, according to (6.21), which in turn reduces the amplitude. So the capacitor C_2 as well as the bias current I, transistor T_1 area and capacitors C_1 and C_2 ratio were optimized in order to achieve good phase noise.

The resonator operates in parallel mode between the series (f_s) and the parallel (f_p) frequencies. The oscillation frequency is given by:

$$f_o = f_s \left[1 + \frac{C_m}{2\,(C_o + C_L)}\right] \tag{6.22}$$

Knowing that the motional capacitance C_m is very small compared to C_o, the oscillation frequency is mainly fixed by f_s as we can note it from (6.22). The oscillation frequency depends also on the loading capacitor C_L, so it could be slightly tuned by using a varactor in series with the resonator. The tuning sensitivity, or the sensitivity of the oscillation frequency to the load capacitance variation, can be expressed as:

$$S = \frac{\partial f_o}{\partial C_L} = -f_r \frac{C_m}{2\,(C_o + C_L)^2} \tag{6.23}$$

Figure 6.19 shows the calculated output power and the phase noise at 100 kHz versus C_1 and for $C_2 = 0.7$ pF. A good trade-off between the maximum output power and the minimum phase noise is achieved with $C_1 = 1.3$ pF.

For the 5 -GHz FBAR resonator used here, the capacitances C_m and C_o are in the order of femtofarad and picofarad, respectively. So the tuning sensitivity is in the

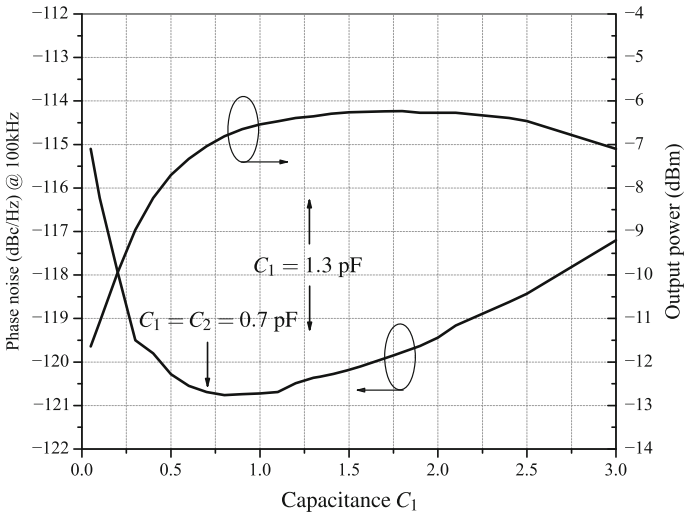

Fig. 6.19 Output power and phase noise @100 kHz versus C_1 ($C_2 = 0.7$ pF)

order of 70 MHz/pF and is not sufficient to accommodate, for example, the whole 200 MHz range of Hiperlan and IEEE 802.11a lower bands [8]. This sensitivity can be trimmed during the design process by adjusting the capacitance C_o via the resonator area, depending on the application needs. As we can see it from (6.23), increasing C_o will reduce the tuning sensitivity. So, if the resonator is used in a fixed frequency oscillator, the capacitor C_o must be large, while it must be low if used in a voltage-controlled oscillator. Moreover, a varactor can be added in series with the resonator in order to compensate for the technological variation of the resonator series resonance frequency f_s. For a 5.32-GHz oscillation frequency, the resonator is dimensioned in order to provide a 5.25-GHz series resonance frequency. The microphotograph of the oscillator is shown in Fig. 6.20, and the chip area is only $640 \times 650\,\mu m^2$.

6.4.2 Balanced Version

The schematic of the balanced oscillator is shown in Fig. 6.21. The balanced output configuration allows driving balanced dividers or mixers, more common than single-ended ones, without the need of a single-ended to differential converter. Without design changes, this configuration consumes twice the DC power of the single-ended one because of the duplicate topology. It also delivers twice the output power and provides better phase noise performance as will be seen in Sect. 6.4.3.

The oscillation frequency and the equivalent resistance of the resonator can be determined by examining the equivalent half circuit of the balanced oscillator shown

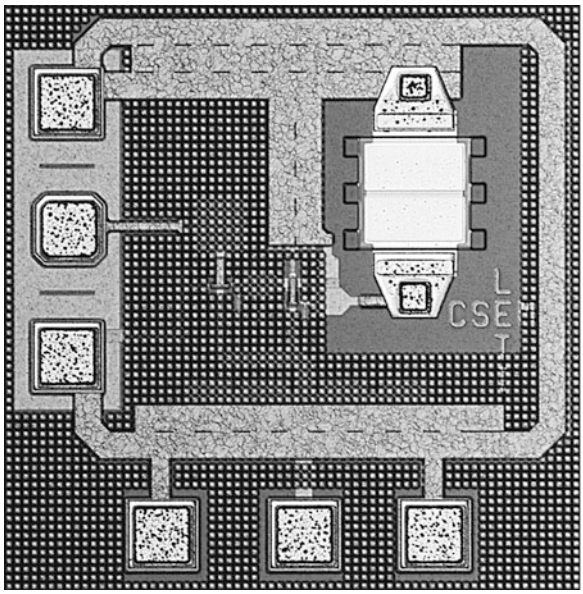

Fig. 6.20 Microphotograph of the single-ended oscillator ($640\times650\,\mu m^2$). ©2006 IEEE. Reprinted, with permission, from [5]

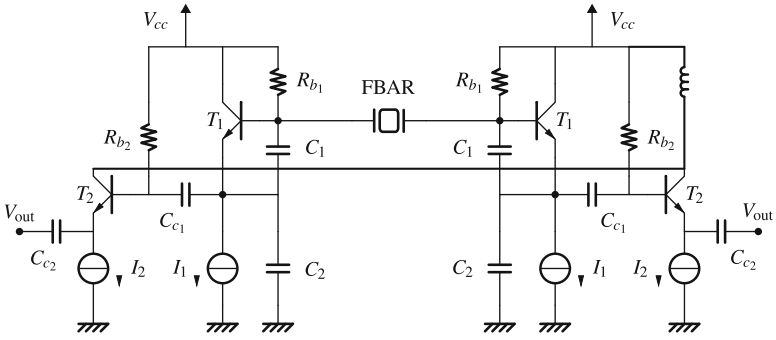

Fig. 6.21 Schematic of the FBAR-based balanced oscillator

in Fig. 6.22. Their expressions are given by the following equations:

$$R_Q = \frac{R_s}{2} + \frac{R_m}{2}\left(1 + \frac{2C_o}{C_L}\right)^2 \tag{6.24}$$

$$f_o = f_s\left[1 + \frac{C_m}{2\left(C_o + \frac{C_L}{2}\right)}\right] \tag{6.25}$$

From (6.24) and (6.25), it is obvious that for the same loading capacitance C_L, the oscillation frequency of this balanced oscillator is higher than the single-ended

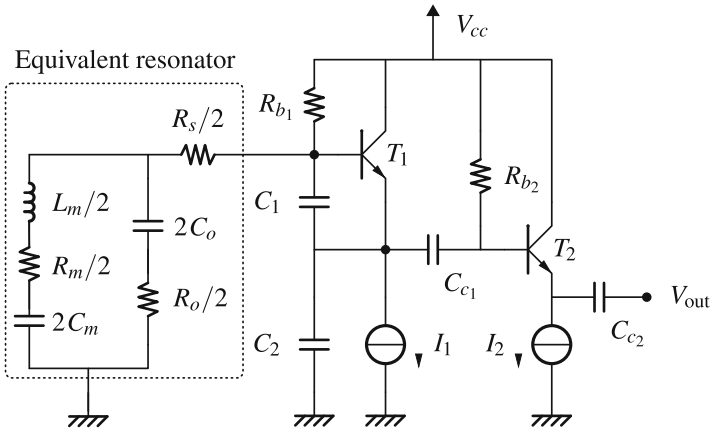

Fig. 6.22 Model of the FBAR-based balanced oscillator

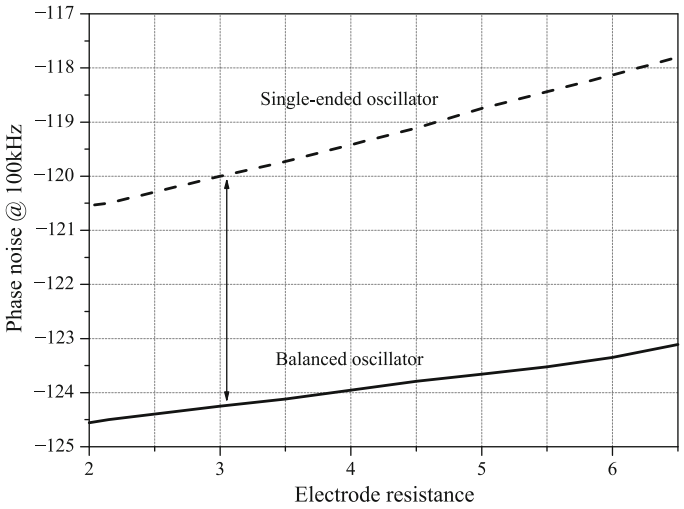

Fig. 6.23 Influence of the electrode resistance on the phase noise of both the balanced and the single-ended oscillators. ©2006 IEEE. Reprinted, with permission, from [6]

one. In this design, the target oscillation frequency f_o is 5.34 GHz, compared to the single-ended one at 5.32 GHz. The resonator equivalent resistance of the balanced oscillator is lower than the single-ended one. Therefore, the balanced oscillator requires less negative resistance, hence less current to start and maintain oscillation. The lower equivalent resistance also implies better phase noise performance. Figure 6.23 shows that the impact of the electrode resistance on the phase noise is less critical for the balanced than for the single-ended oscillator.

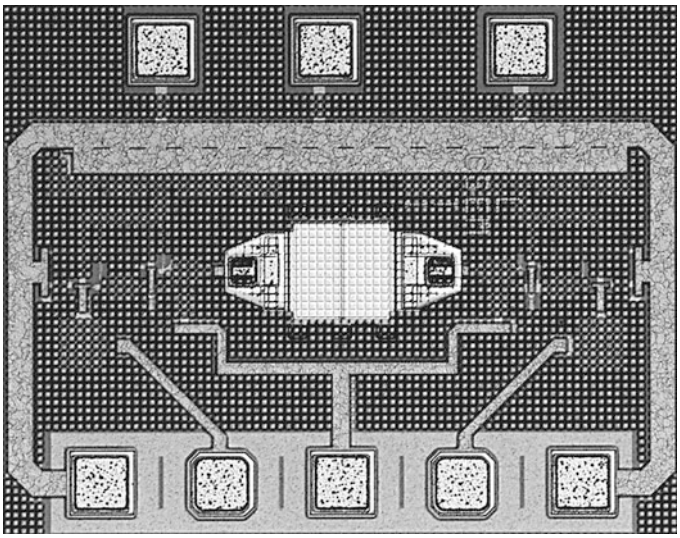

Fig. 6.24 Microphotograph of the balanced oscillator ($830\times650\,\mu m^2$). ©2006 IEEE. Reprinted, with permission, from [6]

Simulations already show that the phase noise is −122.5 dBc/Hz, that is, 4 dB higher than the single-ended version. The microphotograph of the balanced oscillator is shown in Fig. 6.24 and fills $830\times650\,\mu m^2$ of silicon area.

6.4.3 Measurement Results

The single-ended Colpitts core oscillator draws 1.7 mA from 2.7 V supply voltage. The buffer amplifier consumes 3 mA. The balanced Colpitts core draws 2×1.7 mA. On-wafer measurements were done to check the output spectrums and the oscillation frequencies of the oscillators using an RF probe station and a spectrum analyser. The measured oscillation frequency range is between 5.39 and 5.6 GHz. These frequencies are higher than the simulated ones (5.32 GHz for the single-ended and 5.34 GHz for the balanced), due to a resonator resonant frequency higher than anticipated. The power delivered to 50 Ω load is between 8.4 and 11 dBm. Figures 6.25 and 6.26 show the output spectrum of the single-ended oscillator and the balanced oscillator, respectively. The second and the third harmonics' powers are 22 dB and 30 dB below the fundamental, respectively. Phase noise measurements have been carried out on-wafer using the delay-line discriminator method [17]. The measured phase noise of the Colpitts and the balanced oscillators are shown in Fig. 6.27. At 100 kHz offset frequency, the achieved phase noise of the single-ended oscillator is −118 dBc/Hz (simulated value was −118.5 dBc/Hz; the difference is due to a lower quality factor of the resonator). The measured phase noise of the balanced oscillator is −120.5 dBc/Hz at 100 kHz offset from the carrier (simulated

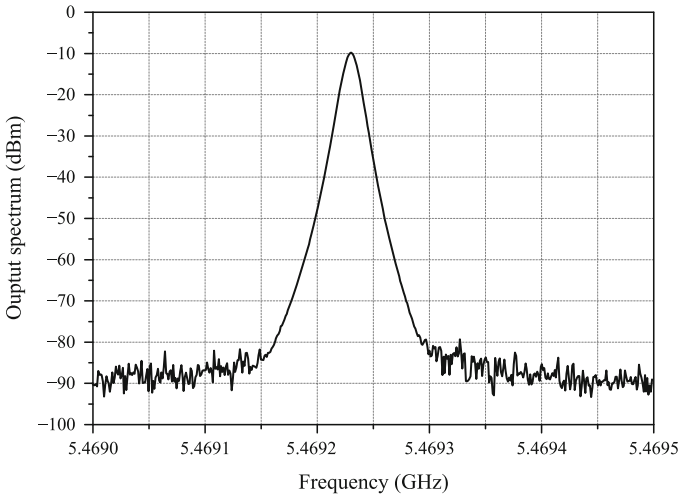

Fig. 6.25 Output spectrum of the single-ended oscillator. ©2006 IEEE. Reprinted, with permission, from [6]

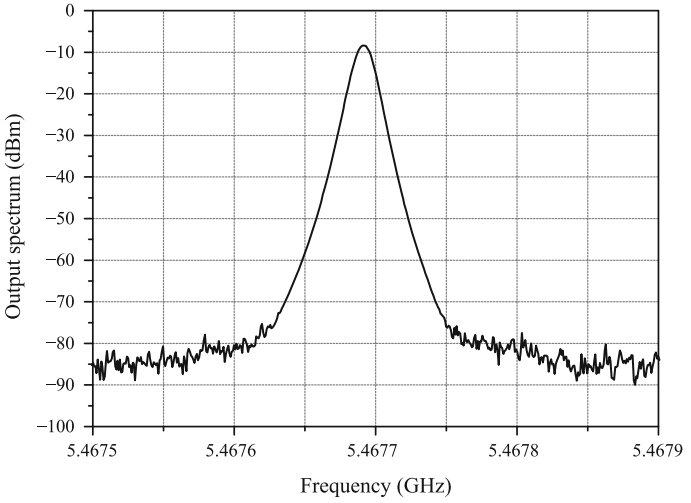

Fig. 6.26 Output spectrum of the balanced oscillator

value was −122.5 dBc/Hz). The figure of merit of both oscillators is around 205.5. The simulated and measured performances of both oscillators are summarized in Table 6.3.

Another advantage of the balanced configuration is that it is well suited for the generation of an oscillation frequency at twice the core oscillator frequency. This also allows doubling the tuning range as well as extending the oscillation frequency. Indeed, oscillators can be designed at higher frequencies by using

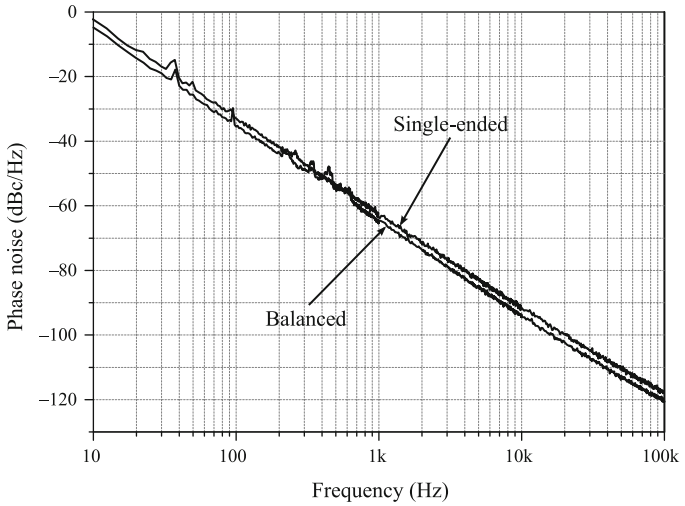

Fig. 6.27 Phase noise of the single-ended and balanced oscillators

Table 6.3 Simulated and measured performances of the two 5-GHz FBAR oscillators

	Colpitts oscillator		Balanced oscillator	
	Simulation	Measure	Simulation	Measure
Power supply V_{cc} (V)	2.7	2.7	2.7	2.7
Current of the core. osc. (mA)	1.7	1.7	2×1.7	2×1.7
Buffer current (mA)	3	3	2×3	2×3
Frequency (GHz)	5.32	5.39; 5.6	5.34	5.4–5.6
Output power (dBm)	−7.5	−11; −8.4	−7.5	−12; −9
Phase noise (dBc/Hz)	−118.5	−118	−122.5	−120.5
Figure of merit (dBc/Hz)	206.4	205.8	207	205.6

resonators with lower resonance frequencies, where it is easier to control the thickness of the piezoelectric materiel, and using multiplication techniques. The generation of the double oscillation frequency can be achieved by shortening the differential output nodes and biasing the collectors of the buffer amplifiers via an inductor as shown in Fig. 6.28.

6.5 LC-Based Balanced Oscillator

To give some elements of comparison, we have also designed two *LC*-based balanced oscillators using STMicroelectronics BiCMOS7RF SiGe:C 0.25 µm, based on standard topologies: a cross-coupled and a Pierce oscillators.

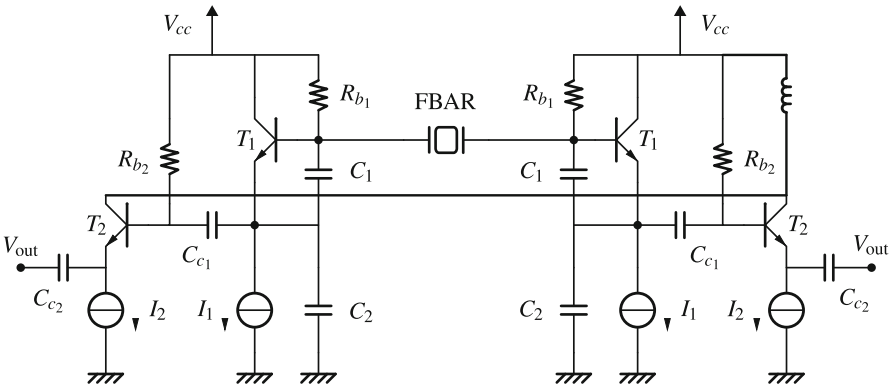

Fig. 6.28 Balanced oscillator with an oscillation frequency at twice the resonator resonance frequency

6.5.1 Cross-Coupled Balanced Oscillator

The schematic of the oscillator is shown in Fig. 6.29. The core of the oscillator consists of a cross-coupled differential pair (T_1, T_2) and the L_cC_v resonator [18]. The capacitive feedback is formed by C_1 and C_π (T_1 or T_2 input capacitance). T_6 and T_7 are buffers that drive 50 Ω loads with an appropriate isolation from the resonator in order to keep its Q-factor as high as possible.

The frequency of oscillation can be estimated as:

$$f_0 = \frac{1}{\sqrt{L_c\, C_{eq}}} \text{ where } C_{eq} = C_p + \frac{C_v C_0}{C_v + C_0} + \frac{C_1 C_\pi}{C_1 + C_\pi} \tag{6.26}$$

C_v is the varactor capacitance, C_p, the parasitic capacitance of the inductor, and C_0 is the output capacitance of the transistors T_1 and T_2. This frequency of oscillation is tuned by changing the P^+/Nwell varactor capacitance with the control voltage V_c. As the substrate is tied to ground, the cathode of such a varactor needs to be tied to AC ground in order to short-circuit the parasitic substrate/Nwell junction capacitor C_{wb} and hence preventing a Q-factor decrease. The varactors are placed in the oscillator as shown in Fig. 6.29 so both sides of the parasitic substrate-Nwell diode are tied to AC ground. The coupling capacitors C_0 are used to isolate the anodes of the varactors from V_{cc}; otherwise, the varactors would be forward biased. These coupling capacitors also help to linearize the capacitance variation with control voltage and thus reduce the phase noise component due to the varactor non-linear characteristics. The drawback of these coupling capacitors is that they reduce the tuning range.

The differential amplitude of the oscillation is set by the tail current I_0 (through T_3) and the equivalent parallel resistance of the LC tank. The oscillation amplitude is given by [19]:

$$V_o \approx \frac{2}{\pi} R_{\text{eq}}\, I_0 \tag{6.27}$$

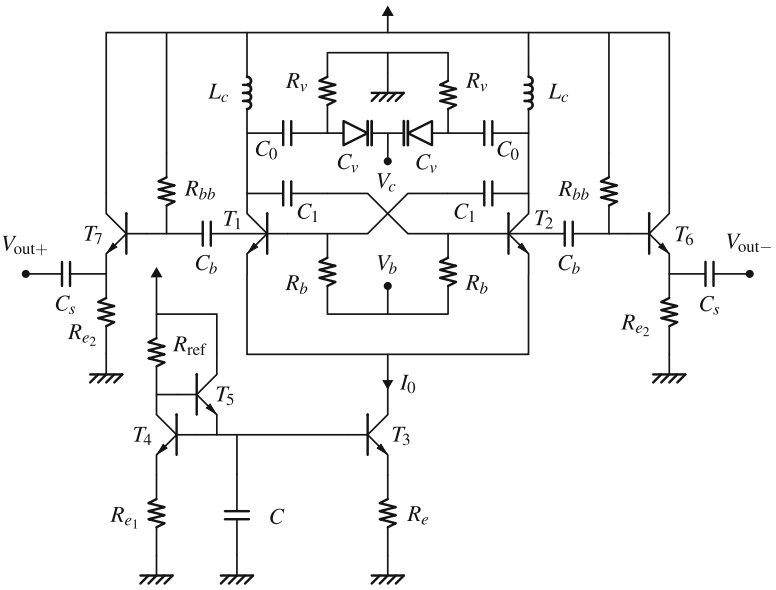

Fig. 6.29 Schematic of the balanced cross-coupled oscillator

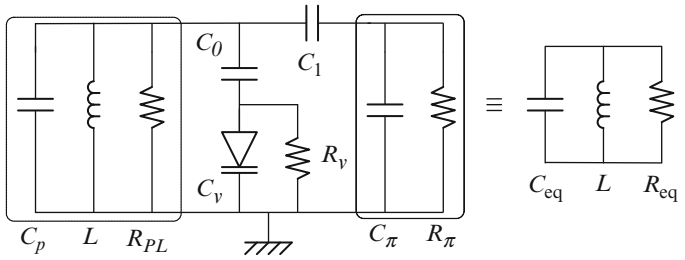

Fig. 6.30 Equivalent resonator from the cross-coupled topology

where R_{eq} is the equivalent parallel resistor of the resonator, which includes the losses in the inductors, capacitors, varactor and the active elements. R_{eq} can be estimated with Fig. 6.30 as:

$$R_{eq} = R_{PL}//R_v\left(1+\frac{C_v}{C_0}\right)^2//\frac{R_\pi}{n^2} \tag{6.28}$$

$R_{PL} = R_s Q_s^2//R_p$ is the equivalent resistance of the inductor, while R_s is its resistance, Q_s its Q-factor, and R_p its parasitic resistance.

The collector-base feedback ratio is:

$$n = \frac{C_1}{C_1 + C_\pi} \tag{6.29}$$

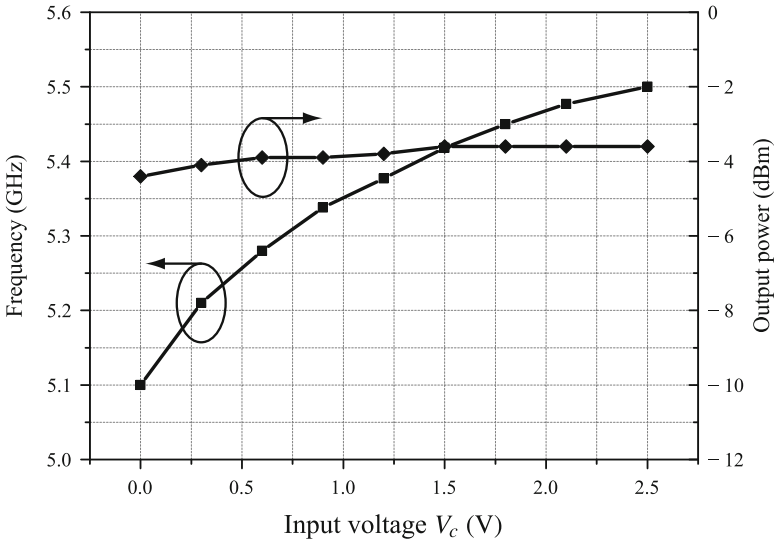

Fig. 6.31 Frequency and output power of the balanced cross-coupled oscillator

Q is the loaded quality factor of the resonator. All the capacitors are implemented as metal-insulator-metal (MIMs) capacitors as they are less lossy (high Q).

In order to reduce the phase noise, the voltage swing of the tank should be as large as possible, but it is limited by the breakdown voltage of the transistors BV_{ce} and the transistors' saturation. If the transistors of the differential pair enter into saturation, the output impedance of the transistors collapses, therefore lowering the quality factor of the resonator leading to a large increase in phase noise.

Therefore, an optimal value of the varactor biasing resistance R_v must be found. A large R_v increases thermal noise, which degrades the phase noise whereas a low R_v deteriorates the quality factor of the resonator, which in turn also increases the phase noise. A good trade-off is a value of 1 kΩ in the present case.

Figure 6.31 shows that the measured output power to a 50 Ω load is about 4 dBm and the range of output frequencies is $[5.1;5.5]$ GHz. Figure 6.32 gives the phase noise at 100 kHz from the carrier, and Fig. 6.33 shows the microphotograph of the balanced cross-coupled oscillator that fills a $945\times1{,}092\,\mu m^2$ silicon area.

6.5.2 Balanced Pierce Oscillator

The schematic of the differential oscillator is shown in Fig. 6.34. The oscillator consists of coupling two identical single-ended Pierce oscillators. The resonator is formed by the inductor L and the series connection of the varactor C_v and MIM

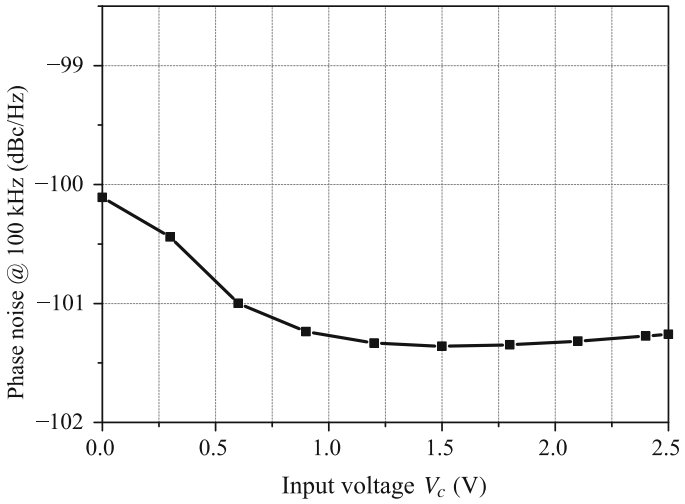

Fig. 6.32 Phase noise of the balanced cross-coupled oscillator

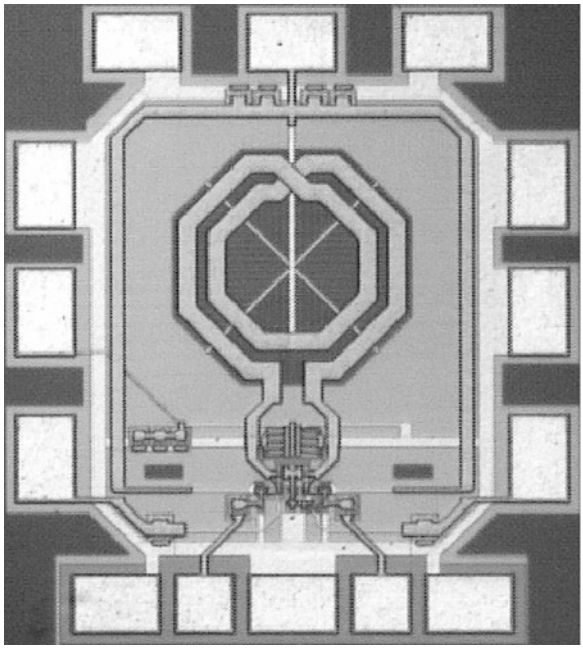

Fig. 6.33 Microphotograph of the balanced cross-coupled oscillator ($945 \times 1{,}092\,\mu m^2$)

capacitor C_0. The feedback ratio, representing the amount of the collector voltage amplitude fed back to the base of the transistor, is given by $n = -C_v/C_b$. Like in the previous balanced cross-coupled oscillator, the feedback ratio n is an important parameter for phase noise minimization.

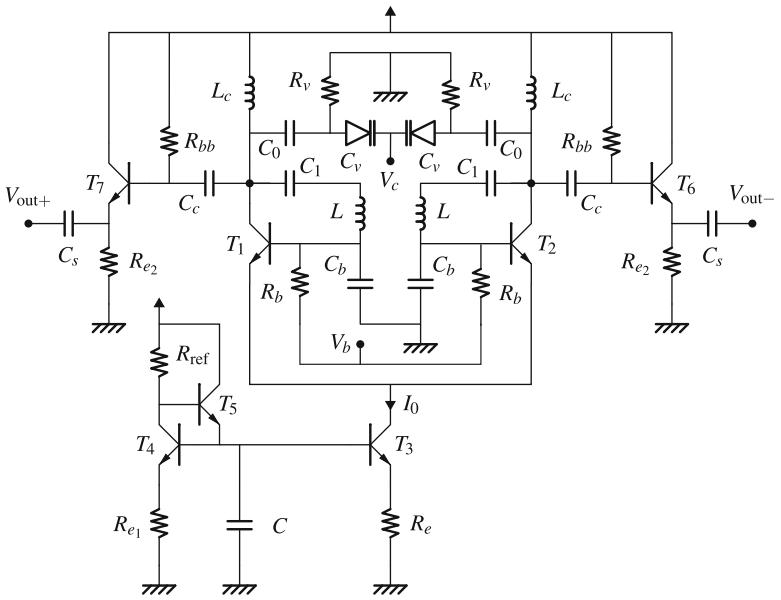

Fig. 6.34 Schematic of the balanced Pierce oscillator

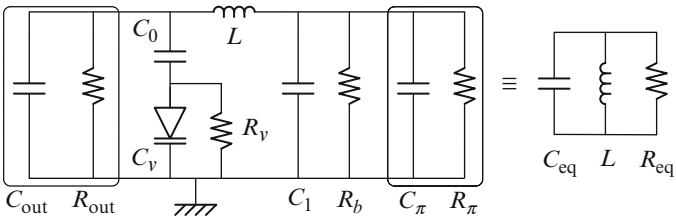

Fig. 6.35 Equivalent resonator from the balanced Pierce topology

The frequency of oscillation is given by:

$$f_0 = \frac{1}{2\pi\sqrt{LC_{eq}}} \text{ where } C_{eq} = \frac{C_v(C_2 + C_\pi)}{C_v + C_2 + C_\pi} \tag{6.30}$$

The resonator amplitude swing is given by:

$$A \approx I_0 R_{eq} \tag{6.31}$$

where I_0 is the tail bias current and R_{eq} is the equivalent parallel resistance of the tank deduced from Fig. 6.35.

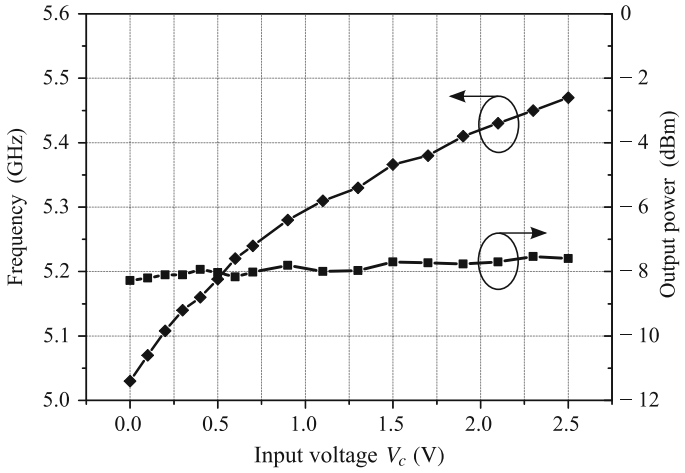

Fig. 6.36 Frequency and output power of the balanced Pierce oscillator

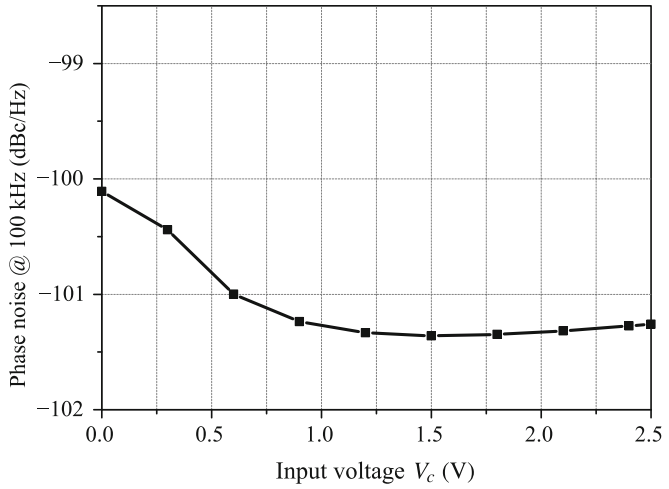

Fig. 6.37 Phase noise of the balanced Pierce oscillator

Figure 6.36 shows that the measured output power to a 50 Ω load is about −8 dBm and the range of output frequencies is [5.05;5.45] GHz. Figure 6.37 gives the phase noise at 100 kHz from the carrier, and Fig. 6.38 shows the microphotograph of the balanced cross-coupled oscillator which fills a 1,000×1,000 μm^2 silicon area.

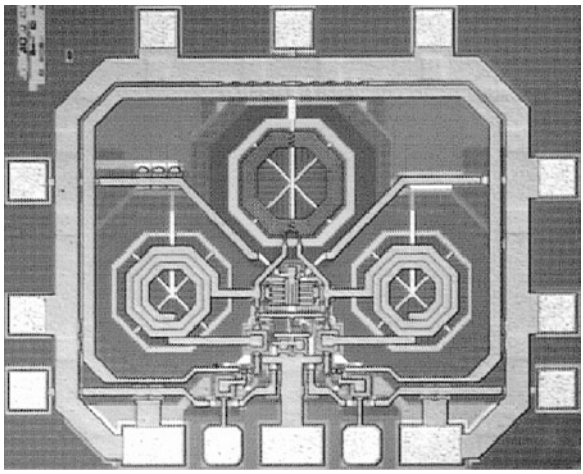

Fig. 6.38 Microphotograph of the balanced Pierce oscillator ($1{,}000 \times 1{,}000\,\mu m^2$)

Table 6.4 LC-based balanced oscillators performances

	Cross-coupled	Pierce
Supply (V)	2.5	2.4
Current (mA)	2.5	3
Tuning range (GHz)	5.10–5.50	5.08–5.50
Output power (dBm)	−4	−8
Core power (mW)	6	7.5
Total power (mW)	32.25	26
Phase noise @ 100 kHz (dBc/Hz)	−99	−101
FoM	−185.5	186.5

6.5.3 LC Based Oscillators Versus FBAR-Based Oscillator

The measured performances of the *LC*-based oscillators are given in Table 6.4. Phase noise and tuning range are very good for *LC*-based oscillators with respect to previously reported data on similar oscillators and fulfil phase noise and bandwidth specifications for IEEE 802.11a. However, a modest 185 FoM is obtained compared to the state-of-the-art 206 FoM demonstrated on our FBAR oscillators (Table 6.3) at 5 GHz. This translates in a 20-dB improvement of the phase noise in the above-IC design which is the consequence of the much higher FBAR *Q*-factor and the low interconnect parasitics.

6.6 Conclusion

In this chapter, we have shown that the above-IC integration of FBAR resonators has allowed the design of low-noise oscillators. Two 5-GHz FBAR-based oscillators were designed using AMI Semiconductor SiGe 0.35-µm BiCMOS process and

above IC integration. A state-of-the-art −120 dBc/Hz phase noise (100 kHz off 5.4 GHz) has been demonstrated. The figure of merit is 205, which is a serious improvement over the best integrated *LC* oscillators that have been investigated in this chapter too. The reduced silicon area needed by the FBAR oscillator allows the realization of small form factor systems. Due to their low tuning range, such oscillators cannot be used as directly as local oscillator for channel selection. However, the oscillator can be used for fully integrated reference signal generation followed by a programmable divider for a completely integrated multimode PLL as well as a fixed RF oscillator for wide-band IF double conversion receivers.

Acknowledgements The author would like to thank all the participants of the MARTINA project and especially AMI Semiconductor (Belgium) for the IC technology, LETI (France) and CSEM (Switzerland) for BAW post-processing.

References

1. K. Lakin, IEEE Microw. Mag. **4**(4), 61 (2003). DOI 10.1109/MMW.2003.1266067
2. Y.H. Chee, A.M. Niknejad, J. Rabaey, in *Proceedings of the Digest of Papers Radio Frequency integrated Circuits (RFIC) Symposium 2005 IEEE*, 2005, pp. 123–126. DOI 10.1109/RFIC.2005.1489606
3. A.P.S. Khanna, E. Gane, T. Chong, in *Proceedings of the IEEE MTT-S International Microwave Symposium Digest*, vol. 2, 2003, pp. 717–720
4. Y. Park, S. Pinkett, J.S. Kenney, W.D. Hunt, in *Proceedings of the IEEE Ultrasonics Symposium*, vol. 1, 2001, pp. 839–842. DOI 10.1109/ULTSYM.2001.991850
5. M. Aissi, E. Tournier, M.A. Dubois, G. Parat, R. Plana, in *Proceedings of the IEEE International Conference Digest of Technical Papers Solid-State Circuits*, 2006, pp. 1228–1235. DOI 10.1109/ISSCC.2006.1696169
6. M. Aissi, E. Tournier, M.A. Dubois, C. Billard, H. Ziad, R. Plana, in *Proceedings of the IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, 2006, p. 4pp. DOI 10.1109/RFIC.2006.1651082
7. M.A. Dubois, J.F. Carpentier, P. Vincent, C. Billard, G. Parat, C. Muller, P. Ancey, P. Conti, IEEE J. Solid State Circ. **41**(1), 7 (2006). DOI 10.1109/JSSC.2005.858627
8. Supplement to ieee standard for information technology telecommunications and information exchange between systems – local and metropolitan area networks – specific requirements. part 11: wireless lan medium access control (mac) and physical layer (phy) specifications: high-speed physical layer in the 5 ghz band, 30 Dec 1999
9. J. van der Tang, D. Kasperkovitz, Proc. ISCAS 2000 Geneva 2000 IEEE Int. Symp. Circ. Syst. 2000 DOI 10.1109/ISCAS.2000.856383
10. M.E. Frerking, *Crystal Oscillator Design and Temperature Compensation* (Van Nostrand, New York, 1978)
11. D.B. Leeson, Proc. IEEE **54**(2), 329 (1966)
12. K. Van Dyke, Proc. IRE **16**(6), 742 (1928)
13. I. Larson, J. D., P.D. Bradley, S. Wartenberg, R.C. Ruby, in *Proceedings of the IEEE Ultrasonics Symposium*, vol. 1, 2000, pp. 863–868. DOI 10.1109/ULTSYM.2000.922679
14. J.F. Rosenbaum, *Bulk Acoustic Wave Theory and Devices* (Arctech House, Norwood, Massachusetts, 1988)
15. R. Ruby, P. Bradley, D. Clark, D. Feld, T. Jamneala, K. Wang, in *Proceedings of the IEEE MTT-S International Microwave Symposium Digest*, vol. 2, 2004, pp. 931–934. DOI 10.1109/MWSYM.2004.1339128

16. K.M. Lakin, in *Proceedings of the IEEE Ultrasonics Symposium*, vol. 2, 1999, pp. 895–906. DOI 10.1109/ULTSYM.1999.849135
17. O. Llopis, J. Juraver, M. Regis, M. Chaubet, J. Graffeuil, Frequency Control Symposium and Exhibition, 2000. Proceedings of the 2000 IEEE/EIA International pp. 511–515 (2000). DOI 10.1109/FREQ.2000.887409
18. H. Jacobsson, S. Gevorgian, M. Mokhtari, B. Hansson, C. Hedenas, T. Lewin, W. Rabe, A. Schuppen, in *Proceedings of the IEEE MTT-S International Microwave Symposium Digest*, vol. 2, 2000, pp. 723–726. DOI 10.1109/MWSYM.2000.863284
19. M. Zannoth, B. Kolb, J. Fenk, R. Weigel, IEEE J. Solid State Circ. **33**(12), 1987 (1998). DOI 10.1109/4.735539

Chapter 7
Low-Power Quadrature Oscillator Design Using BAW Resonators

Shailesh S. Rai and Brian P. Otis

Abstract Aggressive silicon technology scaling, combined with immense unexplored MEMS potential, leads to new advancements in the field of low-power, high-performance RF-MEMS codesigned architectures. This chapter introduces the approach, theory, and design of an important block in RF transceivers: a quadrature voltage-controlled oscillator (QVCO) stabilized with bulk acoustic wave (BAW) resonators. A new time-varying source degeneration coupling mechanism has been used to quadrature-couple the two oscillator cores for I/Q signal generation, reducing the required headroom of the series coupling transistors. The phase accuracy of the quadrature outputs determines the image rejection capability of integrated RF front ends. To improve phase error in the presence of process mismatch, we introduce a self-calibration loop for phase control of quadrature oscillators. Design concerns like temperature stability of BAW-tuned oscillators are addressed. Side-by-side comparisons of the BAW- and LC-based oscillators are made, and assembly challenges are discussed.

7.1 Introduction

CMOS processes suffer from the unavailability of high-quality factor passives, which results in severe performance limitations of RF oscillators. With continuous scaling of CMOS technology nodes, the cost of integrating large on-chip passives like inductors becomes a big concern. Addressing these limitations, RF designers have begun to explore alternatives to make RF systems power, bandwidth, and cost efficient. One approach is the use of high-quality factor MEMS-assisted RF design. In this chapter, we discuss design approaches for one of the most critical blocks

S.S. Rai (✉) • B.P. Otis
University of Washington, Seattle, USA
e-mail: shailesh@u.washington.edu; botis@u.washington.edu

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_7,

in communication systems: the quadrature voltage-controlled oscillator (QVCO). We demonstrate a 600-μW bulk acoustic wave (BAW)-based QVCO prototype, providing a figure of merit (FOM) significantly greater than the state-of-the-art. An identical QVCO using an integrated LC tank with 600-μW power consumption is presented for direct comparison with the BAW-tuned QVCO.

Digital RF transceiver architectures increasingly require quadrature (I/Q) local oscillator sinusoid generation to allow highly integrated image-reject or direct-conversion architectures. The phase noise and power consumption of QVCOs have a significant impact on the system performance. The relevance of VCO power consumption becomes even more profound due to the demand for implementations of ultralow-power, short-range, wireless transceivers. For last decade or so, the technique of quadrature signal generation using two cross-coupled on-chip LC VCOs has been extensively explored.

The next section discusses the need for quadrature signals and the potential of high-Q MEMS components in RF systems. A brief overview of existing QVCO design techniques using two-core coupling is also presented in Sect. 7.2. Section 7.3 introduces to BAW technology and discusses approaches for BAW-assisted RF design. A BAW-based QVCO prototype is presented in Sect. 7.4, along with an autocalibration loop for QVCO phase-error correction.

7.2 Quadrature Modulation, RF-MEMS Potential, and QVCO Design Background

In this section, we discuss the need for quadrature signals in modern communication systems and how they have been affecting the modulation approaches over the past few years. Also, the potential of MEMS-assisted RF design is presented, followed by a brief overview of published QVCO design techniques using on-chip passives as a tuning element.

7.2.1 Quadrature Signals in RF Communication Systems

By definition, a quadrature signal is a complex signal consisting of two orthogonal and noninterfering signals exactly 90° out of phase. From RF system point of view, we need to ask: *What purpose in-phase (I) and in-quadrature (Q) signals serve, and why are they required in modern communication systems?* The answer lies in the type of modulation employed in modern communication systems. The push towards minimizing power while maintaining high spectral efficiency (bps/Hz) has resulted in the ubiquity of digital (de)modulation schemes. Digital modulation techniques offer better efficiency, performance, and adaptability than analog modulation techniques. Most digital modulation schemes rely on the quadrature modulation of I/Q

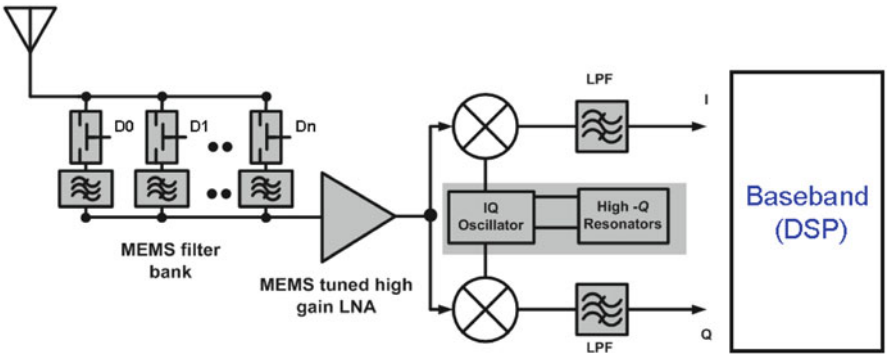

Fig. 7.1 High-Q passive enhanced front end

signals, allowing single-sideband (SSB) modulation. Zero-IF (direct-conversion) receivers rely on quadrature downconversion, and fully integrated image-reject downconverters necessitate precise I/Q sinusoid generation.

The I/Q phase and amplitude matching becomes critical in determining quadrature modulation accuracy and efficiency. Upconverter image suppression is highly dependent upon the phase and amplitude imbalance between I and Q signals. The relation between I/Q amplitude mismatch, phase mismatch, and image rejection, IR, in dB is [1]:

$$IR = 10log\frac{1+2(1+\delta)cos\varepsilon+(1+\delta)^2}{1-2(1+\delta)cos\varepsilon+(1+\delta)^2}, \tag{7.1}$$

where ε is the I/Q phase offset from 90° and δ is the I/Q amplitude mismatch in dB. Although specifications vary widely, it is typically necessary to limit I/Q phase error to less than a few degrees. In communication systems, the specification of adjacent channel interference rejection defines the I/Q imbalance requirement. For example, 42 dB IR needs less than 0.1 dB I/Q amplitude mismatch and less than 1° I/Q phase offset error from 90°.

7.2.2 Need for MEMS in RF Systems

The power/noise trade-off in wireless systems is a well-known issue and has been extensively studied and documented in the past few decades. The use of on-chip passive components consuming a large silicon area adds a third dimension in defining the RF systems: the cost metric.

Over the past few years, research groups have begun focusing on RF design using high-Q MEMS components (RF switches, varactors, resonators, and filters). Shown in Fig. 7.1 is a possible example of how MEMS components can be used at various stages in RF front end. A very low loss with narrow bandpass response

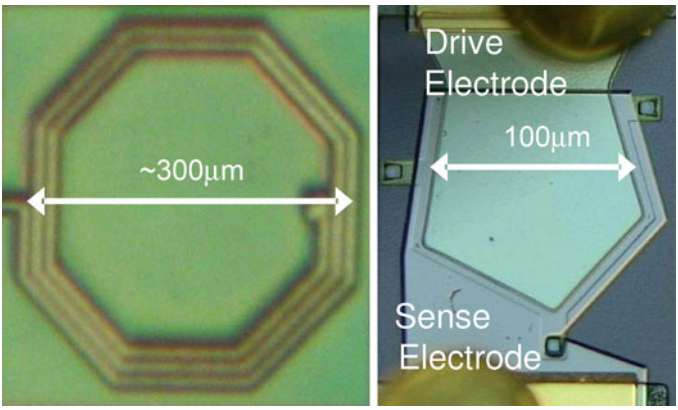

Fig. 7.2 Area comparison of on-chip inductor ($Q = 10$), *left*, and BAW resonator ($Q = 1{,}000$), *right*

MEMS filter bank can be used at the receiver input for interferer rejection, allowing a power reduction in high-gain LNAs by relaxing the linearity specification. MEMS resonators can be used as tuning element in RF frequency synthesis, which is the emphasis of this chapter. Also, the future integration of MEMS processes with CMOS will result in high level of integration that will address cost issues for communication systems. MEMS switches would be ideal for highly reconfigurable RF systems because of their extremely low loss characteristics and high RF power handling capabilities. Similarly, BAW resonators with high unloaded quality factor ($>1{,}000$) find its application as a tuning element in ultralow-power oscillator design. Oscillators using high quality factor (Q) BAW resonators have been demonstrated to provide excellent phase noise, supply pushing, and power consumption performance.

On-chip LC VCOs exhibit a well-known power consumption and phase noise trade-off [2]. Additionally, the silicon area consumed by on-chip inductors is becoming increasingly expensive in terms of lost transistor count and functionality. An area comparison for a typical on-chip inductor (Q approximately 10) and BAW resonator (Q approximately 1,000) is shown in Fig. 7.2.

It is striking to note that, in a 65-nm process, a single $300 \times 300\,\mu$m inductor consumes the real estate of approximately one million logic or memory transistors. The fact that designers are willing to trade this amount of functionality for the ability to place one inductor with a poor quality factor on chip indicates how important tuning elements are in RF design.

To improve the trade-off between oscillator power consumption and phase noise while reducing wasted die area, we developed a prototype that demonstrates the use of matched high-Q BAW resonators used as a tuning element in a 2-GHz QVCO. We introduce a new oscillator coupling mechanism for quadrature signal generation, which we demonstrate on both a BAW- and an integrated LC-tuned QVCO [3].

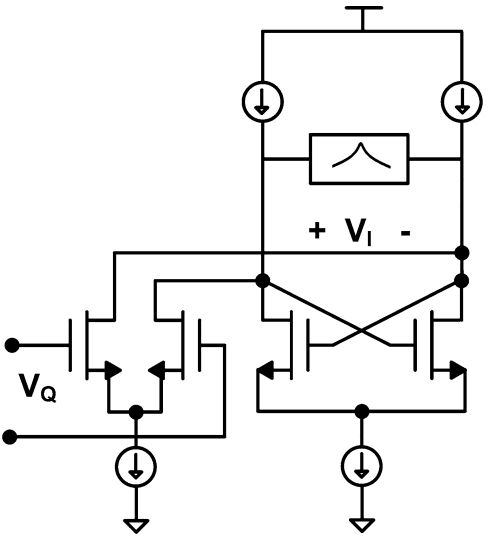

Fig. 7.3 Quadrature signal generation through anti-phase signal injection (©2008 IEEE. Reprinted, with permission, from [3])

7.2.3 QVCO Design Approach

Does the world need another quadrature signal generator? Recent literature suggests that, especially in low-power space, quadrature LO signal generation consumes a relatively large amount of system power budget. Many circuits have been proposed for generating quadrature LO signals. One common technique is to utilize a VCO operating at twice the desired LO frequency followed by a quadrature divider. However, this technique has the drawback of higher power and area consumption. Another typical method involves a VCO followed by an RC-CR phase-shifting network. Quadrature accuracy for this method is determined by the absolute accuracy of on-chip passive elements and results in a loss of LO power through the passive structure. One of the most power-efficient solutions is the antiphase cross coupling of two symmetric differential LC oscillators, first demonstrated by Rofougaran et al. [4]. As seen in Fig. 7.3, if V_Q couples in a signal at the same frequency as the free-running oscillator frequency, the coupling results in shifting the phase of the output signal V_I, as V_Q modifies the vector current summation at the output node [5].

V_I and V_Q are generated using identical oscillators cross-coupled in antiphase. Quadrature oscillators can be intuitively understood by observing that antiphase coupling enforces a 180° phase shift across the two oscillators, naturally leading to a 90° phase in the intermediate node. Alternately, the in-phase coupling of two identical oscillators simply results in in-phase coupling, which is not of interest in this application.

The following quadrature oscillator coupling methods have been demonstrated in literature:

1. Parallel-coupled QVCO (PQVCO), in which the coupling transistors are placed in parallel with $-g_m$ switching transistors [4]. As shown in [6], this coupling mechanism suffers from a trade-off between phase noise and phase error of quadrature signals.
2. Series-coupled QVCO, in which the coupling transistors are placed in series with the negative-gm switching transistors [6]. Although this coupling mechanism results in better phase noise than the PQVCO, the need for stacking saturated transistors invariably increases the device headroom. Series coupling has also been demonstrated with linear-region devices stacked below the switching transistors [7].
3. Back-gate-coupled QVCO, in which the two oscillators are coupled using the back gate of $-g_m$ switching transistors [8]. The back-gate coupling mechanism necessitates triple-well CMOS process availability and introduces the possibility of forward biasing the intrinsic bulk-substrate diode.
4. Other methods include inductively coupling the common source node of two oscillators [9] and transformer coupling [10], both of which require extra die area resulting from using additional inductors.

As explained later in Sect. 7.4.2, we have attempted to address these limitations by introducing oscillator coupling based on time-varying source degeneration of cross-coupled RF transistors in differential oscillator.

7.3 BAW Technology and MEMS-Assisted RF Design

7.3.1 BAW Resonator as Tuning Element in RF Design

In this subsection, an overview of BAW technology and its scope for use in integrated circuit design is presented. Figure 7.4(a) shows the cross section of a thin-film bulk acoustic wave (FBAR or BAW) resonator.

The BAW design consists of a freestanding membrane containing a piezoelectric material, aluminum nitride (AlN), sandwiched between two molybdenum electrodes [11]. The resonance mechanism is based on the energy exchange between the mechanical and electrical domains. The electrical signal induces a mechanical displacement in the piezoelectric film, generating acoustic waves. In turn, this mechanical displacement itself induces an electric field, creating a resonant condition at a particular frequency. The high Q results from high acoustic isolation due to an air interface on both sides of the resonator, resulting in very little acoustic power dissipation during the membrane motion at resonance.

One of the properties of BAW resonator that is of interest for radio frequency circuit designers is the much higher quality factor (Q) available than any other

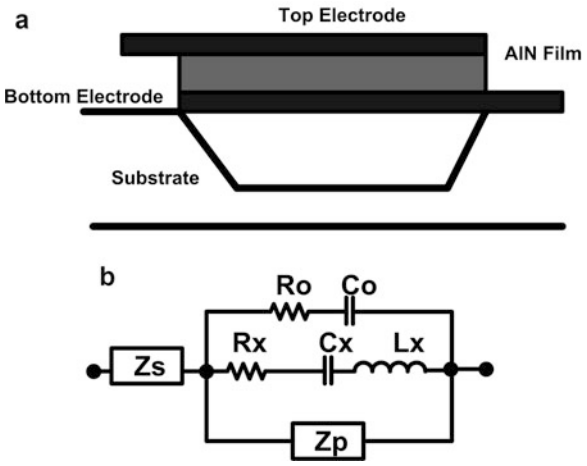

Fig. 7.4 (**a**) Cross section of BAW resonator and (**b**) BAW equivalent electrical model for simulation (©2008 IEEE. Reprinted, with permission, from [3])

on-chip resonant structure. For the purposes of circuit co-simulation of BAW resonators using standard tools like SpectreRF, the fundamental resonance mode can be modeled using the modified Butterworth–Van Dyke model (MBVD), as shown in Fig. 7.4b, where C_0 represents the parallel plate capacitance of the BAW resonator [12]. As shown in (7.2), the oscillation frequency of an oscillator operating at the parallel resonance is [13]:

$$f_{osc} = f_s \cdot \sqrt{1 + \frac{C_x}{C_T}}, \tag{7.2}$$

where f_s is the series resonant frequency of a BAW resonator and is modeled by L_x and C_x, and C_T is the total capacitance accounting for C_0 and any capacitive loading from CMOS circuitry. The series and parallel loading effects of CMOS circuitry on the BAW resonator are represented by Z_s and Z_p, respectively, with R_x and R_0 representing resonator losses.

Though BAW resonators typically have an unloaded Q greater than 1,000 for resonators fabricated in the 0.5–7.5 GHz frequency range, the real impedance at the parallel resonance is around 1 kΩ. In other words, the extremely high Q for BAW resonators does not directly translate into a corresponding higher value of parallel resistance as an inductor would, avoiding the need for an extremely large circuit impedance to avoid detuning the resonator. This can be further understood with the aid of a simulated plot of impedance versus frequency for a BAW resonator as well as for an on-chip LC tank (Q≈10), shown in Fig. 7.5.

Even though both resonant tanks are tuned to approximately 2.1 GHz, there are significant differences in the impedance profile. At low frequencies, the BAW resonator is capacitive and exhibits a high impedance, while the on-chip LC tank is inductive at low frequencies and allows DC current flow. The LC tank exhibits one low-Q resonance with a parallel resistance of approximately 800 Ω. At its series resonance, the BAW resonator impedance is real with a value less than 5 Ω, while at

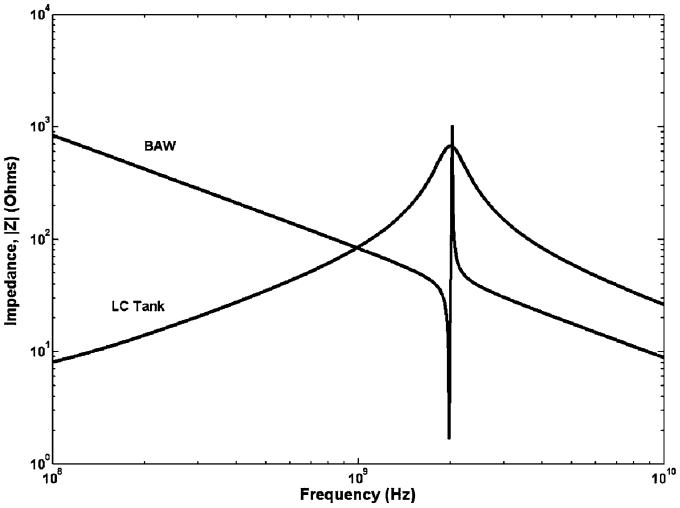

Fig. 7.5 Simulated impedance magnitude versus frequency for BAW and LC tank. The maximum impedance of BAW at parallel resonance is around 1 kΩ (©2008 IEEE. Reprinted, with permission, from [3])

the parallel resonance, the impedance is real with a value greater than 1 kΩ. From RF designers' perspective, we can visualize BAW resonator as a tuning element that replaces on-chip inductor for frequency tuning of RF oscillators.

RF oscillator design using these high-Q BAW resonators mainly suffers from the following disadvantages:

1. *Low tuning range*: Because of their extremely high quality factor, the tuning range of BAW-tuned oscillators is extremely narrow. Current research involves techniques like widening the tuning range of BAW-tuned VCOs using negative capacitance [14]. Ultimately, there is a direct trade-off between phase noise and tuning range for BAW-tuned RF oscillators.
2. *Integration*: Even though BAW resonators have extremely small form factor, which bodes well for the integration of MEMS with CMOS processes and miniaturizing RF systems, they are most often implemented off chip and present assembly challenges.
3. *Temperature stability*: The BAW resonators have a temperature coefficient of approximately -30 ppm/°C. This issue can be addressed using either on-chip temperature compensation techniques or by exploring ways to perform mechanic temperature compensation.

Similar to frequency stabilization efforts that are commonly used for quartz crystal references, we can address temperature instability using an on-chip real-time temperature calibration of BAW-tuned oscillator. A conceptual example is shown in Fig. 7.6. The on-chip temperature calibration using a discrete switched capacitor array is performed as a two-step process:

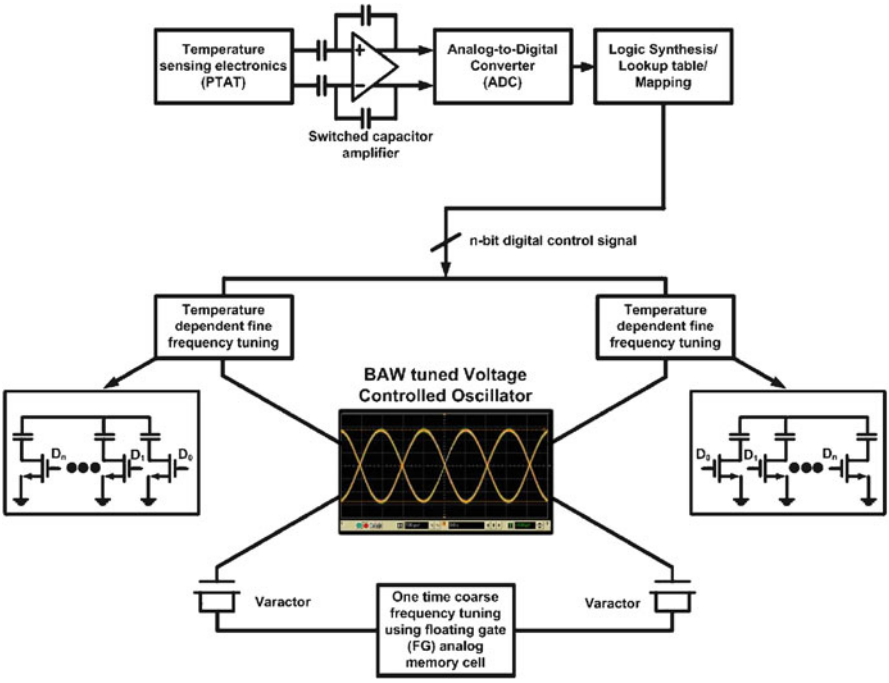

Fig. 7.6 Proposed real-time temperature calibration of BAW-tuned oscillator

1. *Sensing, amplifying, and digitizing temperature variations*: This can be accomplished by a combination of PTAT temperature sensor, a high-gain switched capacitor amplifier and a high-resolution ADC.
2. *Fine frequency tuning using a lookup table (LUT)*: A prestored temperature compensation profile can be stored in an on-chip SRAM. The ADC addresses the LUT as the temperature changes, resulting in an appropriate frequency correction capacitor setting.

Research in this space is aimed at eventually providing thin-film quartz-free frequency references. In addition to temperature drift, a calibration step must correct any manufacturing variations. As shown in Fig. 7.6, calibration can be accomplished using standard MOS varactors, while storing the calibration data in on-chip nonvolatile memory (e.g., floating gate (FG) analog memory cells).

A quartz-free RF frequency reference would exhibit a phase-noise profile characteristic of a free-running oscillator. Figure 7.7 shows the measured phase noise for a differential 1.5-GHz GPS-band 300-μW oscillator.

The far-out phase noise is better than −150 dBc/Hz at 1-MHz offset. Since the close-in noise is not locked to a low-frequency reference, the integrated jitter can suffer. Tackling the frequency stability and the close-in phase-noise challenges of BAW-based references is a topic of current research [15].

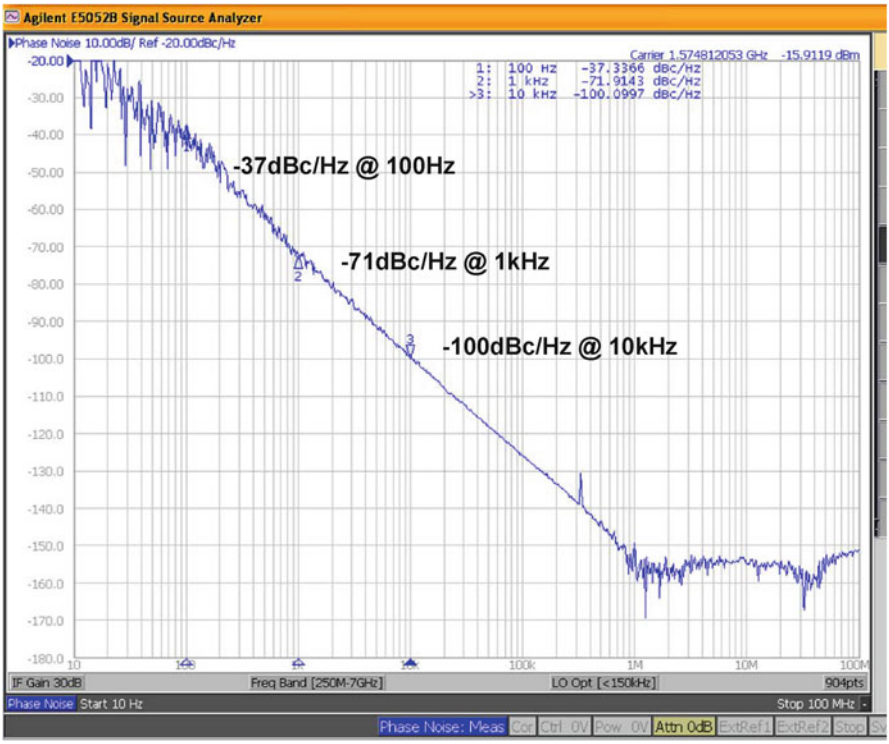

Fig. 7.7 Measured phase noise of BAW-tuned oscillator at GPS frequency

7.4 BAW-Tuned QVCO Analysis

This section presents the analysis and implementation of a BAW-based QVCO (BQVCO). BQVCO design is a two-step process:

1. CMOS/BAW codesign of differential oscillator core.
2. The coupling of two symmetrical oscillator cores for quadrature signal generation.

The design focus is optimization for ultralow power consumption, I/Q quadrature accuracy, and low phase noise. These circuit techniques are also applicable for design of RF QVCOs using on-chip passives. Additionally, a phase correction loop is proposed for autocalibration of I/Q phase imbalance.

7.4.1 RF Oscillator Design Using BAW Resonators

The differential oscillator circuit topology must drive the BAW resonator symmetrically and excite its parallel resonance while maintaining a sufficiently high

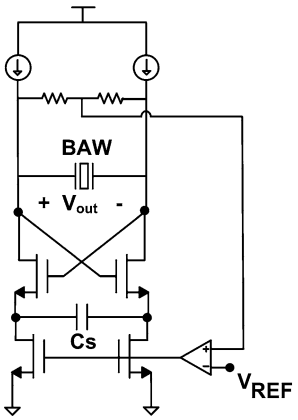

Fig. 7.8 BAW-based differential oscillator (©2008 IEEE. Reprinted, with permission, from [3])

impedance at resonance to avoid detuning the mechanical resonator. In many traditional integrated differential oscillators, the negative resistance of the sustaining amplifier is achieved through a cross-coupled transistor pair. Looking into the cross-coupled pair, a wideband, DC-to-RF negative resistance is created. To achieve oscillation at a specific frequency, a parallel LC load is used to provide high impedance at one frequency, as shown in Fig. 7.5. At this particular frequency, the total tank impedance is negative and oscillation occurs. The bias current can conveniently be provided through the same inductors that tune out the tank capacitance. A BAW resonator-based differential oscillator requires a circuit topology, satisfying the following requirements:

1. *Low-frequency stability*: The resonator can be shunted across the cross-coupled pair, providing a high-Q response at the parallel resonance to set the oscillation frequency. However, at low frequencies, the resonator presents a high impedance (as shown in Fig. 7.5). Thus, the circuit would be DC unstable if presented with a broadband negative resistance.
2. *Loaded Q*: Bias current must be supplied to the cross-coupled pair without detuning the BAW resonator at RF frequencies. Thus, a high-impedance differential circuit environment at high frequencies is required.
3. *Common-mode control*: A fully differential high-impedance bias circuit necessitates common-mode control over the oscillator core.

Current sources could be used to supply bias current to the cross-coupled pair without detuning the resonator, but would result in low-frequency instability. One way to circumvent this problem is to design a high-pass response into the cross-coupled pair negative resistance. This is realized by using separate current sources for the cross-coupled pair and coupling the sources through a capacitor [16]. At low frequencies, the cross-coupled pair experiences a large degeneration impedance, reducing the negative resistance. At high frequencies, the sources of cross-coupled pair interact, providing full transconductance from the cross-coupled pair. The schematic of a BAW-based differential oscillator core is shown in Fig. 7.8.

It can be shown that the differential impedance looking into the cross-coupled pair is given by

$$Z_{cc} = -\frac{2}{g_m}\left(1 + \frac{g_m}{2sC_s}\right). \tag{7.3}$$

Thus, at high frequencies, the structure provides a negative resistance of $-2/g_m$. If C_s is large, low-frequency instability is possible due to the inductive nature of the capacitively degenerated cross-coupled pair interacting with the BAW parallel plate capacitance. This resonance would cause a parasitic oscillation if the loop gain exceeded 0 dB following Barkhausen's criterion, which could be attenuated by reducing C_s. However, reducing C_s increases the negative resistance pole frequency in (7.3), reducing the oscillator loop gain at the desired resonance. Thus, a proper choice of C_s is crucial for stable and efficient oscillation. The use of BAW resonators in place of an LC tank necessitates a common-mode feedback circuit, which was implemented using resistive common-mode sensing and a single-stage operational transconductance amplifier (OTA). To avoid detuning the BAW resonator, the sensing resistor should be significantly larger than the effective resonator impedance at the parallel resonant frequency (shown in Fig. 7.5). The common-mode voltage is adjustable by changing the reference bias voltage of the OTA.

7.4.2 Coupling Mechanism and BQVCO Topology

As explained in Sect. 7.2.3, the different coupling mechanisms reported in literature exhibit multiple trade-offs. We propose to address those issues using a circuit topology employing a new coupling mechanism. Figure 7.9 shows the proposed BAW QVCO architecture with two differential oscillator cores coupled in antiphase for quadrature signal generation.

The voltage headroom and parasitics added by the quadrature coupling transistors should be minimized to allow low supply voltage and maximum tuning range, respectively. We propose the use of differential triode-region switching transistors placed in series with the source decoupling capacitor (C_s). These transistors, driven by opposing oscillator core, degenerate the cross-coupled pair, effectively modulating their negative transconductance. In absence of any signal injection from the second core, the effective small-signal transconductance, G_M, is approximately given by $g_m/(1 + g_m/g_{ds})$, where g_m is the intrinsic transconductance of cross-coupled switching transistors and $1/g_{ds}$ represents the on-resistance of coupling transistor operating in the linear region. $1/g_{ds}$ is a time-varying parameter in this design, as an oscillating signal is applied to the coupling transistors for quadrature coupling of two oscillator cores. As a result, the cross-coupled oscillator transconductance is time-varying and phase-locked to the opposing oscillator core. Compared to conventional QVCO coupling methods, this coupling mechanism consumes no voltage headroom enabling operation at low power supplies. For stable oscillation, the transconductance of each cross-coupled switching transistors must

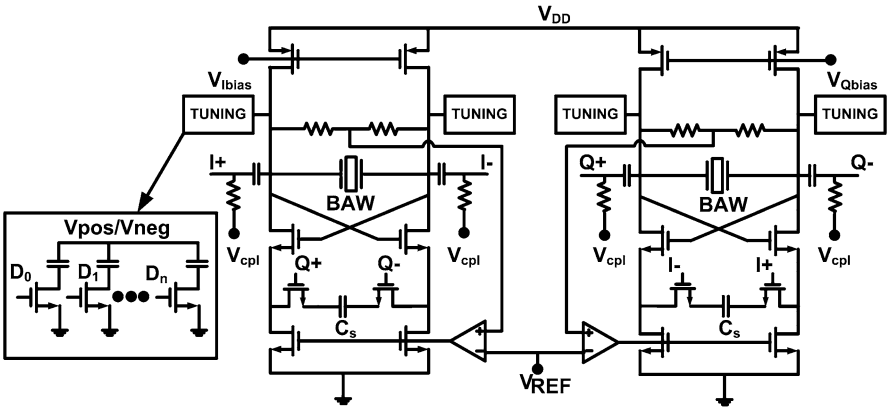

Fig. 7.9 BAW-based quadrature voltage-controlled oscillator (©2008 IEEE. Reprinted, with permission, from [3])

be greater than $2/R_p$. The sizing of the switching devices presents a trade-off between transconductance efficiency and device f_T degradation. In this work, the cross-coupled transistors are sized for moderate inversion operation with an inversion coefficient (IC) of approximately 0.3. Future technology nodes will allow RF circuits biased into increasingly weak levels of inversion, allowing higher transconductance efficiencies for a given device f_T [17]. Coarse frequency tuning of the QVCO is achieved using a digitally controllable capacitor bank at each output node. The MOS switches controlling discrete capacitor switching has been optimized to avoid detuning of BAW resonator.

7.4.3 Comparison of BAW- and LC-Stabilized Oscillators

Both QVCOs were implemented in a 0.13-μm 8-metal CMOS process. A micrograph of the assembled QVCO chip is shown in Fig. 7.10.

The CMOS chip is wirebonded to two frequency-matched BAW die. An LC QVCO using an identical oscillator topology is placed on the same die for direct comparison, which utilizes two on-chip 7-nH inductors with unloaded Q of approximately ten in place of the BAW resonators. Figure 7.11 shows the measured quadrature signals generated with the BQVCO.

The coupling transistor bias is set to minimize I/Q phase error. The measured I/Q phase error for both the BAW- and LC-based QVCOs is less than $\pm 1^\circ$, as limited by the test setup. A comparison of the measured frequency spectrum with 1-MHz span and resolution bandwidth of 5 kHz for the BAW and LC QVCOs, both operating at a 600-μW power consumption and 1-V power supply, is shown in Fig. 7.12. The significantly higher unloaded Q provided by the BAW resonators (approximately

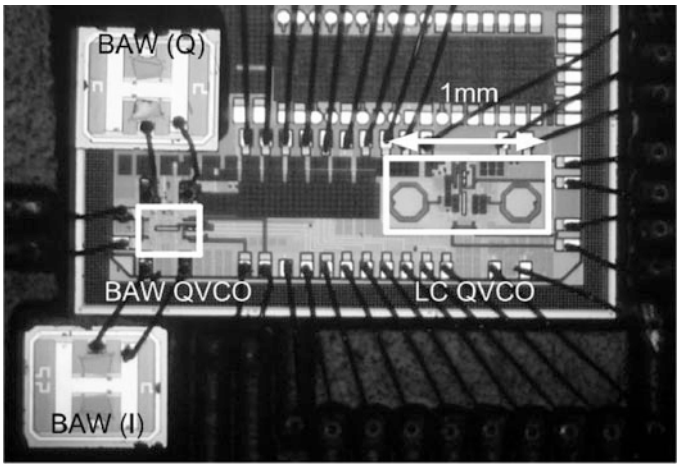

Fig. 7.10 Chip micrograph of BAW and LC QVCO (©2008 IEEE. Reprinted, with permission, from [3])

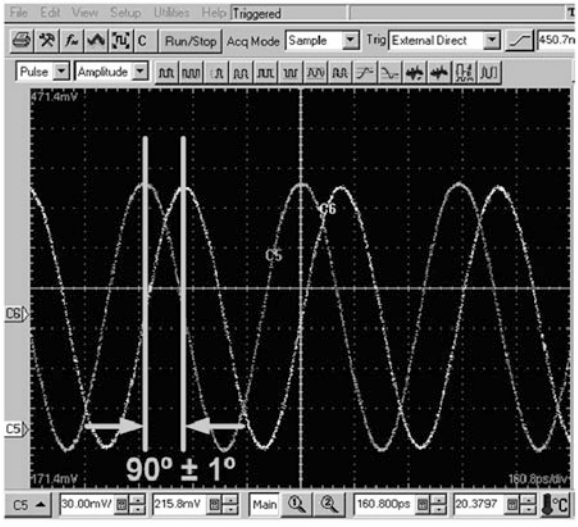

Fig. 7.11 Measured BAW QVCO quadrature sinusoid output (©2008 IEEE. Reprinted, with permission, from [3])

1,500) compared to the integrated LC tanks (approximately 10) clearly improves the quality of the frequency spectrum and thus the phase noise.

The high Q mechanical resonance, however, greatly reduces the tuning range of the BQVCO to 1.5 MHz. The sensitivity to capacitance variation for the BAW and LC oscillators is given by (7.4) and (7.5), respectively:

$$\frac{\delta f_{osc}}{\delta C_T} = f_{series} \frac{-C_x}{2C_T^2}, \tag{7.4}$$

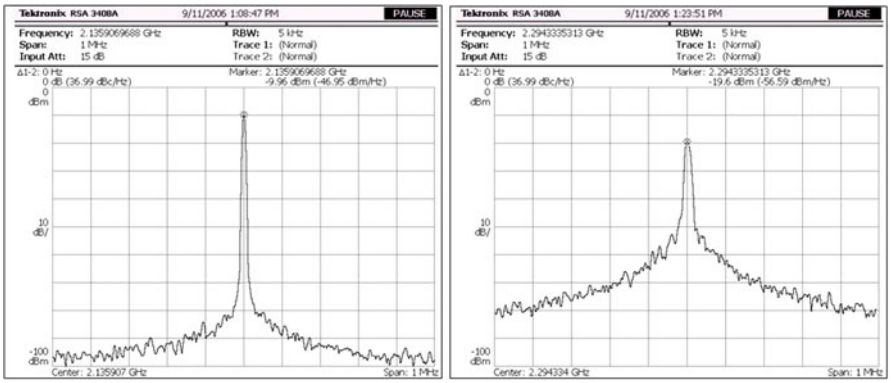

Fig. 7.12 Measured BAW- and LC- QVCO spectra (©2008 IEEE. Reprinted, with permission, from [3])

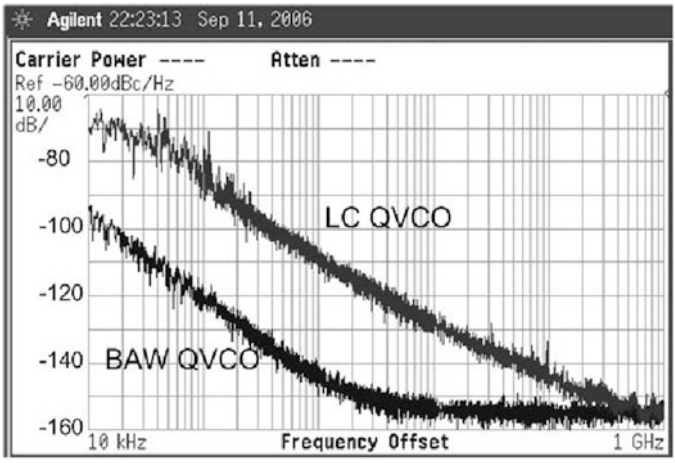

Fig. 7.13 Measured BAW- and LC-QVCO phase noise (©2008 IEEE. Reprinted, with permission, from [3])

$$\frac{\delta f_{osc}}{\delta C} = \frac{-C^{-3/2}}{4\pi\sqrt{L}}. \tag{7.5}$$

While severely limiting the tuning range, this insensitivity to change in capacitance results in the ability to perform fine-grained discrete frequency control with relatively large switched capacitors. The LSB of the capacitor array corresponds to 60-kHz (30 ppm) change in carrier frequency for BQVCO, with a frequency sensitivity of −0.006 MHz/fF. The LC QVCO achieves a tuning range of approximately 300 MHz, with a frequency sensitivity of −1.6 MHz/fF. The measured phase noise of both oscillators operating at 600 μW, overlaid for comparison, is provided in Fig. 7.13.

Table 7.1 Prototype BAW and LC QVCO specifications

	BAW QVCO	LC QVCO
Technology	0.13 µm CMOS	0.13 µm CMOS
Center frequency	2.1 GHz	2.2 GHz
Tuning range	<1%	13.6%
V_{dd}	1 V	1 V
Power consumption	600 µW	600 µW
Phase noise @ 1 MHz	−143.5 dBc/Hz	−110.7 dBc/Hz
FOM	212.1 dB	179.7 dB
Active CMOS area	290 × 360 µm	440 × 1,040 µm

The phase noise for the BQVCO is −143.5 dBc/Hz at 1-MHz offset from the 2.1-GHz carrier frequency. The LC QVCO achieves a phase noise of −110.7 dBc/Hz at 1-MHz offset from the 2.2-GHz carrier. Although the design value is 600 µW, the BQVCO achieves reliable startup at a minimum power consumption of 250 µW from a 1-V power supply. The standard FOM for oscillators is given by the following:

$$FOM = 10log\left[\left(\frac{f_o}{\Delta f}\right)^2 \frac{1}{L(\Delta f) \times P_{diss}\,|_{mW}}\right]. \tag{7.6}$$

The performance summary for the BQVCO and LC QVCO is presented in Table 7.1. The BQVCO operating at its nominal 600-µW operating point achieves a FOM of 212.1 dB, which exceeds any previous published QVCO implementation. The FOM of the LC QVCO operating at 600 µW at a 1-V power supply is 179.7 dB, which is comparable to other published LC QVCOs.

7.4.4 Quadrature Autocalibration Loop

As explained in Sect. 7.2.1, I/Q amplitude and phase matching is the key to defining quadrature modulator efficiency in RF systems. Amplitude control loops for RF oscillators have been researched and well documented. One of the fundamental problems with quadrature-coupled VCOs is the error in the 90° phase relation. Quadrature imbalance has multiple negative effects in a digital transceiver, including a reduction in image rejection and degradation of signal-to-noise ratio at the detector. For a quadrature VCO using the two-core coupling implementation, the phase accuracy is mainly determined by frequency matching between the resonant tanks, parasitic capacitances, device mismatches, and mutual coupling between inductors. For on-chip LC-tuned QVCOs, inductive coupling between tanks has been shown to introduce systematic phase error [18]. BAW-tuned QVCOs, on the other hand, exhibit very little tank coupling. However, frequency mismatch between two high Q resonant tanks in BAW-based cross-coupled oscillators can result in I/Q phase offset error.

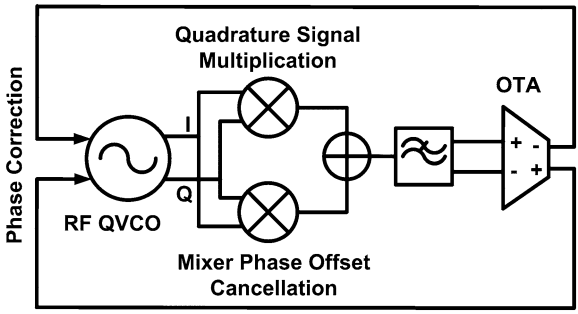

Fig. 7.14 Proposed phase calibration architecture for RF QVCO

Shown in Fig. 7.14, we propose a new quadrature correction scheme that will provide a large improvement in quadrature accuracy for any RF QVCO. Our technique involves an analog feedback loop that drives the phase error to zero.

The feedback point is the differential control of RF varactors that controls the frequency of the two oscillators. Differential adjustments of the varactor control voltage modify the natural resonant frequency of the oscillator tanks and thus modulate the phase relation between the oscillator outputs. To detect the phase relation, we make use of the following trigonometric identity:

$$sin(\omega_1 t + \theta_\varepsilon) \cdot cos(\omega_2 t) = \frac{1}{2}\left[sin((\omega_1 + \omega_2)t + \theta_\varepsilon) + sin((\omega_1 - \omega_2)t + \theta_\varepsilon)\right], \quad (7.7)$$

where the sine and cosine terms represent the I and Q outputs of the QVCO and θ_ε is the error of the phase relation at the output of the QVCO. If the output of the operation is low-pass filtered, the output becomes

$$\frac{1}{2}\left[sin((\omega_1 - \omega_2)t + \theta_\varepsilon)\right]. \quad (7.8)$$

Since the two oscillators are frequency locked, $\omega_1 = \omega_2$. Thus, the output becomes

$$\frac{1}{2}\left[sin(\theta_\varepsilon)\right], \quad (7.9)$$

which provides a zero DC output with a zero phase error.

The differential I and Q outputs of the QVCO are mixed by a pair of double-balanced Gilbert-cell mixers. Since the I/Q inputs exhibit different phase shifts through the mixer, phase imbalance in the signal path is introduced. As shown in Fig. 7.15, two mixers can be used to cancel the phase offset that exists between the upper and lower inputs of the traditional Gilbert cell. The outputs of the mixer are added in the current domain and are subsequently low-pass filtered, leaving a fully differential DC term that is proportional to the phase error. A fully differential OTA is used to drive the phase error to zero by feeding back to differential varactors controlling the QVCO core oscillation frequency. Simulation results suggest that this technique can reduce a 5° phase error (induced by tank mismatch) to less than 0.5°.

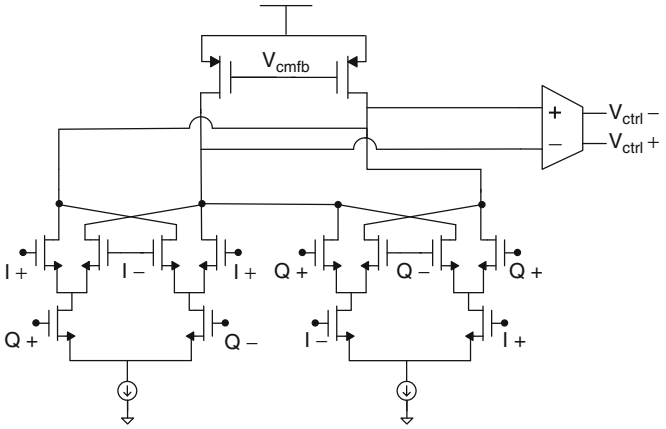

Fig. 7.15 Quadrature mixer using two Gilbert cells

7.5 Conclusion

BAW resonator-tuned quadrature oscillators offer compelling power and noise performance improvements over traditional LC-tuned oscillators. A new time-varying source-degenerated oscillator coupling mechanism is proposed for quadrature signal generation. The combination of a high-Q mechanical resonator and source-degenerated quadrature coupling technique mitigates the oscillator power/phase noise trade-off and reduces the necessary circuit headroom. The same circuit architecture was demonstrated with an on-chip LC tank for comparison. Eliminating the need for on-chip inductors significantly reduces the necessary die area at the expense of additional assembly complexity.

The use of MEMS components into RF circuits has resulted in systems with superior performance. As RF designers, we need to continually explore the immense potential of high quality factor MEMS passives in communication systems. A very important step will be inclusion of MEMS passive models in design libraries of commercial CMOS processes. Extensive ongoing research in the field of MEMS-assisted RF designs will address design of other critical RF blocks and will revolutionize our existing wireless communication systems.

Acknowledgements The authors would like to acknowledge Avago Technologies for the BAW resonator fabrication and Richard Kim for his careful layout efforts.

References

1. Q. Gu, *RF System Design of Transceivers for Wireless Communications* (Springer, New York, 2005)
2. T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits* (Cambridge University Press, Cambridge, 2004)

3. S. Rai, B. Otis, IEEE J. Solid State Circ. **43**(1), 300 (2008)
4. A. Rofougaran, J. Rael, M. Rofougaran, A. Abidi, *IEEE ISSCC Digest of Technical Papers*, 1996, pp. 392–393
5. B. Razavi, *Design of Integrated Circuits for Optical Communications* (McGraw-Hill Science/Engineering/Math, New York, 2003)
6. P. Andreani, A. Bonfanti, L. Romano, C. Samori, IEEE J. Solid State Circ. **37**(12), 1737 (2002)
7. J. Chang, Y. Youn, M. Park, C. Kim, *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, 2003, pp. 295–298
8. H. Kim, C. Cha, S. Oh, M. Yang, S. Lee, IEEE J. Solid State Circ. **39**(6), 952 (2004)
9. S. Gierkink, S. Levantino, R. Frye, C. Samori, V. Boccuzzi, IEEE J. Solid State Circ. **38**(7), 1148 (2003)
10. A. Ng, H. Luong, *IEEE ISSCC Digest of Technical Papers*, 2006, pp. 711–720
11. R. Ruby, P. Bradley, J. Larson III, Y. Oshmyansky, D. Figueredo, *IEEE ISSCC Digest of Technical Papers*, 2001, pp. 120–1
12. J. Larson III, P. Bradley, S. Wartenberg, R. Ruby, IEEE Ultrason. Symp. **1** (2000)
13. B. Otis, J. Rabaey, IEEE J. Solid State Circ. **38**(7), 1271 (2003)
14. P. Vincent, J. David, I. Burciu, J. Prouvee, C. Billard, C. Fuchs, G. Parat, E. Defoucaud, A. Reinhardt, *IEEE ISSCC Digest of Technical Papers*, 2008, pp. 478–629
15. S. Rai, Y. Su, A. Dobos, R. Kim, R. Ruby, W. Pang, B. Otis, *IEEE Frequency Control Symposium*, 2009, pp. 385–387
16. D. Ruffieux, *Proceedings of ESSCIRC*, 2002, pp. 85–88
17. N. Pletcher, J. Rabaey, *Proceedings of ESSCIRC*, 2005, pp. 387–390
18. A. Mazzanti, F. Svelto, P. Andreani, IEEE J. Solid State Circ. **41**(6), 1305 (2006)

Chapter 8
Tunable BAW Filters

Stéphane Razafimandimby, Cyrille Tilhac, Andreia Cathelin, and Andreas Kaiser

Abstract Most current radio receivers use SAW bandpass filters for band selection. Their center frequency depends on the dimensions of the interdigital structure. Therefore, in the next few years, SAW technology will face some fabrication process limits because of the need of higher operating frequencies. BAW resonators are emerging as an alternative technology. One of the potential advantages is the compatibility of the BAW process with standard silicon processing technology. BAW resonators exhibit high Q factors around 1,000 which make them very attractive for low insertion loss RF filters. Nevertheless, process dispersions on the thickness of AlN lead to a shift of the BAW resonator characteristic frequencies and thus a shift of the filter's center frequency. Moreover, BAW resonators suffer from thermal drift of around 20 ppm/°C. Therefore, the need of designing a tunable filter and an automatic tuning circuitry appears for SoC in order to correct process and/or temperature deviations. After presenting the way of synthesizing BAW filters for a given mask, the feasibility of tuning a 2-GHz bandpass BAW filter is demonstrated in this chapter by a very first flip-chip assembly of a BAW solidly mounted resonator (SMR) die on the top of a BiCMOS 0.25-µm chip.

S. Razafimandimby (✉) • C. Tilhac • A. Cathelin
STMicroeletronics, Crolles, France
e-mail: stephane.razafimandimby@st.com; cyrille.tilhac@st.com; andreia.cathelin@st.com

A. Kaiser
ISEN/IEMN, Lille/Villeneuve d'Ascq, France
e-mail: andreas.kaiser@isen.fr

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_8,

8.1 Tunable BAW Filter Synthesis and Its Physical Implementation

8.1.1 BAW Resonator

BAW resonators are typically composed of three parts: electrodes, a piezoelectric layer, and an isolation part like an air gap for TFBAR (thin-film bulk acoustic resonator) or a Bragg reflector for SMR (solidly mounted resonator) (see Fig. 8.1). The principle of the isolation part is that a change in the impedance affects the amount of acoustic energy that is reflected and transmitted. Creating a discontinuity at material boundaries allows breaking the transmission of an acoustic wave in the materials. Hence, a Bragg mirror consists of several pairs of alternatively high and low acoustic impedance $\lambda/4$ material layers. Thus, most of the signal is reflected in the piezoelectric material and not transmitted to the substrate.

8.1.2 Filter Topologies

8.1.2.1 Ladder Filter

In this chapter, we denote the BAW resonator's impedance in the series branch by R_s and the BAW resonator's impedance in the parallel branch by R_p. As BAW resonators behave like a short circuit at f_s and like an open circuit at f_p, the series frequency of R_s is aligned to the parallel frequency of R_p in order to create a bandpass filter function. For this, a loading material is added on the top of R_p to reduce its resonant frequencies as illustrated on the SMR of Fig. 8.2. In the case of a ladder filter, f_s of R_p and f_p of R_s create the notch frequencies. In fact, R_p at f_s forms a RF path to ground, and R_s at f_p cuts the RF signal transmission. If R_s and R_p have the same size, that is, the same C_o, each cascaded stage brings 6-dB attenuation out of the band (see Fig. 8.2).

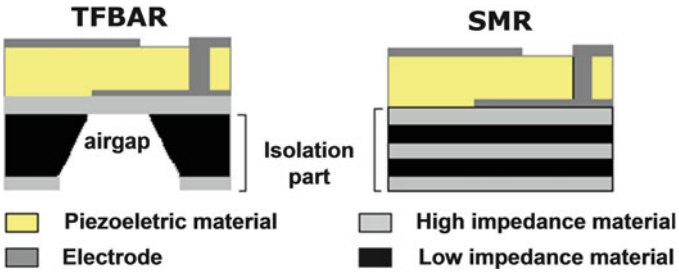

Fig. 8.1 Different types of BAW resonators

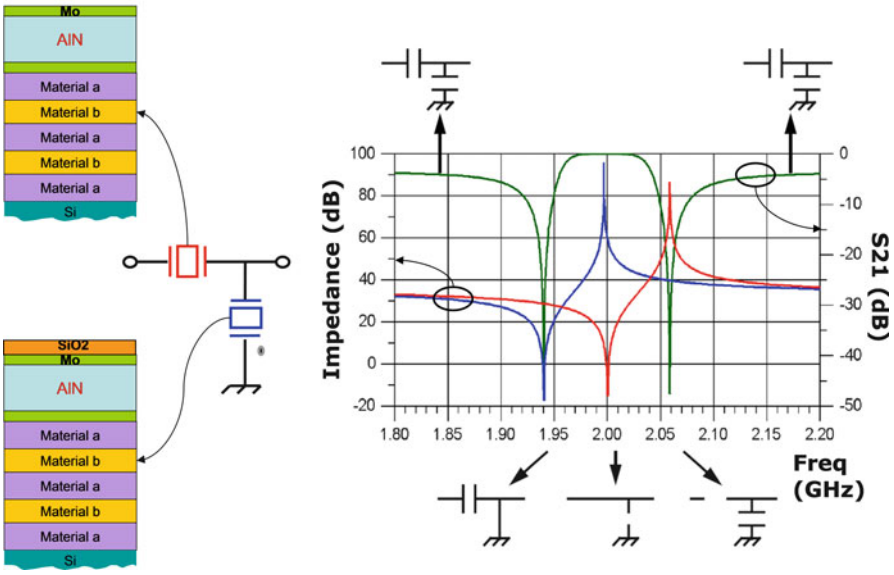

Fig. 8.2 Ladder filter principle

8.1.2.2 Lattice Filter

The lattice filter can be analyzed in the same way as the ladder filter. No notch frequency is produced because of the existence of a perpetual RF path. However, a second means of using lattice filter can be explained by analyzing the transfer function of such a structure (see Fig. 8.3a). The filter transfer function is defined by

$$S_{21} = \frac{R_0(R_s - R_p)}{(R_s + R_0)(R_p + R_0)}. \tag{8.1}$$

When series and parallel arm impedances are equal and their phases are opposite, we obtain an optimal condition to transmit the RF signal (see Fig. 8.3a). Contrary to that, when the impedances of the different branches are equal in magnitude and phase, a high attenuation is obtained. In fact, this new way of understanding lattice filters is based on a phase constructive phenomenon (see Fig. 8.3b). Considering BAW lattice filter, only one BAW resonance is thus useful. The low-impedance resonant frequency f_s is therefore preferred, and the antiresonance is no more useful.

To tune this type of filters, we have to tune the resonant frequency of each resonant structure unless one of the resonant frequencies can be made no more useful. Notice that a common control quantity between R_s and R_p can make the tunability easier. Finally, lattice filters exhibit a better selectivity than ladder filters. Therefore, for a same given filter mask, lattice structures employ a lesser number of BAW resonators. Furthermore, the differential structure eliminates the constraints on even-order nonlinearity.

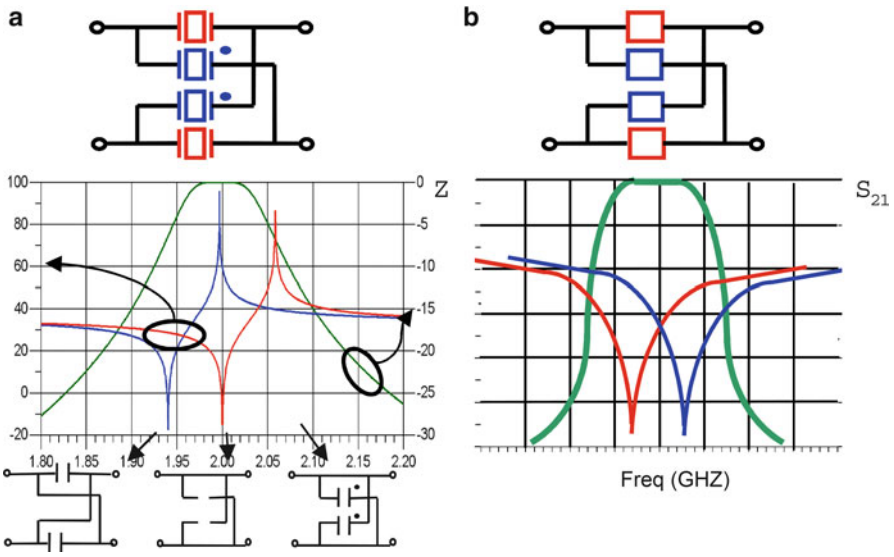

Fig. 8.3 Lattice filters principle. (**a**) Classical one. (**b**) Constructive phenomenon

8.1.3 Filter Synthesis

The technique used to synthesize a tunable BAW filter is based on a classical design technique for polynomial LC filters in which the coupling concept is used.

In order to realize a tunable BAW filter, the first step is to define or to choose a synthesis method that allows introducing the electrical model of the BAW resonator. A brief review of classical filter synthesis exhibits two theories: the image parameters theory and the effective parameters theory. Only the second one seems to be useful. While the starting point of the first method is the effective attenuation (directly linked to the insertion losses) which characterizes the behavior of the network, the effective parameters allow more freedom in the architecture, which in our case seems sensible, given the nature of the architecture to be realized. Indeed, the main goal is to synthesize a tunable filter using BAW resonators which have a motionless electrical model. Finally, the results obtained by the image parameters theory can be likened to a particular case of those provided by the effective parameters theory because, for a set of given parameters (ripple, bandwidth…), a unique architecture is possible, contrary to the multitude of architectures obtained with the effective parameters. It turns out that the method of the effective parameters is more rigorous than the image parameters method because it defines the effective attenuation exactly.

The transfer function of the chosen filter can be defined by several categories of functions, of which the most common are Butterworth, Tchebychev, and generalized Tchebychev, which will be not detailed in this chapter.

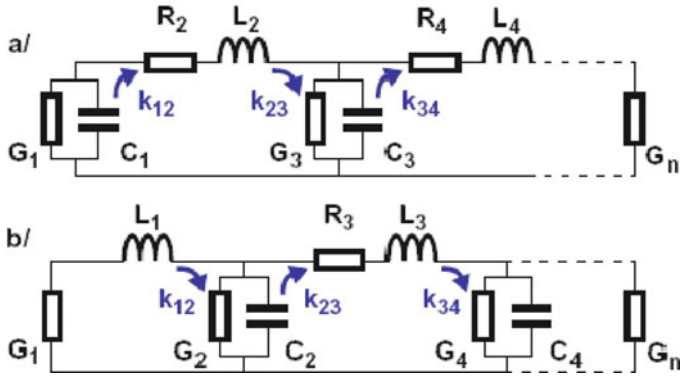

Fig. 8.4 Principle of coupling method with (**a**) a capacitor as first element and with (**b**) an inductor as first element

The filter synthesis is generally based on an equivalent electrical schematic of the device with lumped elements. This synthesis allows, from the circuits theory, to determine the values of the components according to the chosen filtering function. In order to synthesize some filters (lowpass, bandpass, highpass…), a lowpass prototype is necessary at the beginning. This one is defined thanks to its own lossless transfer function. We then use classical transformation principle to obtain the equivalent bandpass filter. Despite a very simple method, it is impossible to create a BAW filter with good performance because of the use of inductors with critical Q in the direct signal path.

The bandpass filter synthesis by the coupling method introduced by Milton Dishal in the networks synthesis theory [1, 2] has therefore been preferred. The word coupling generally refers to a parasitic effect between an element to another one (see Fig. 8.4).

The two coefficients used in this method are defined as

$$q_i = \frac{\omega_{3dB} L_i}{R_i} = \frac{\omega_{3dB} C_i}{G_i}, \tag{8.2}$$

$$k_{ij} = \frac{\omega_{ij}}{\omega_{3dB}} = \frac{1}{\omega_{3dB}\sqrt{L_{ij} C_{ji}}}, \tag{8.3}$$

where q_i corresponds to the quality coefficient of the reactive element i and k_{ij} is the coupling coefficient which is actually equal to the ratio between the pulsation ω_{ij} of the i and j elements and the filter cutoff pulsation ω_{3dB}.

The bandpass filter consists of a coupling between series or shunt resonators. As the BAW electrical model is based on series LC resonator, we will take the coupled series resonator arrangement.

The obtained ladder configuration consists of a series resonator coupled through inductive or capacitive inverters, networks. The negative elements of the inverter will be integrated in the BAW model in the next part. In fact, the k coefficient can be

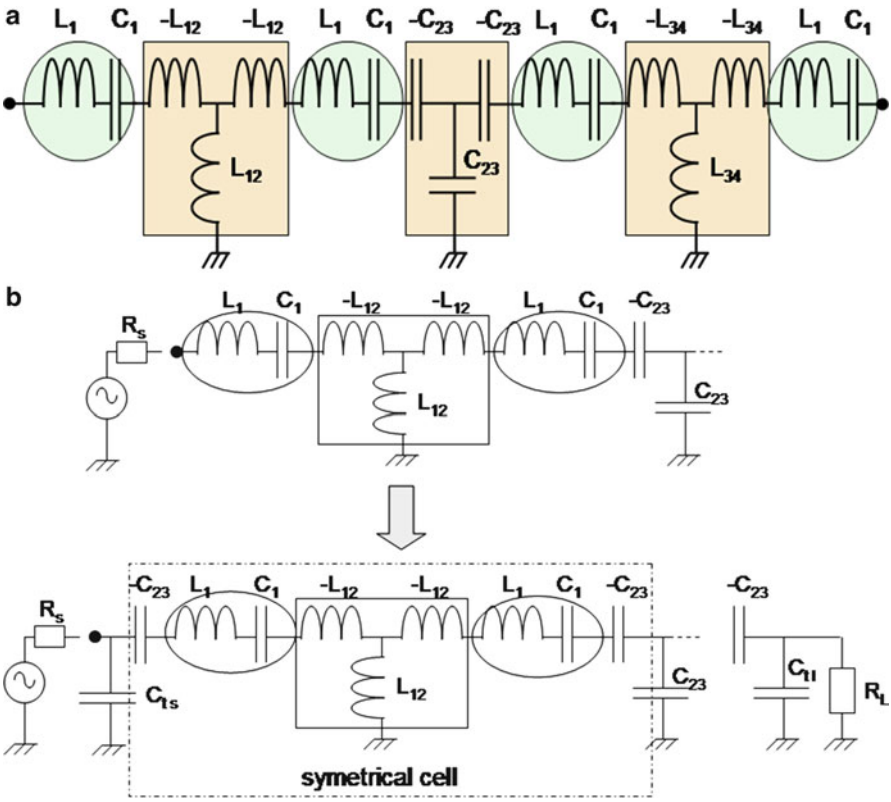

Fig. 8.5 (**a**) 4th-order ladder filter. (**b**) Transformation to a BAW ladder filter

identified to a coupling coefficient between two adjacent resonators, whereas the q coefficient represents the input–output adaptation. The bandwidth is strongly linked to the coefficient k, while the ripple depends on k and q.

The impedance and admittance inverters are ideal coupling networks that have the interesting property to be able to change their impedance without changing the central frequency of the filter. It is thus possible to adjust the filter bandwidth by changing the inverter or the values of its elements [3].

Let's see the different steps implemented for a bandpass filter developed for a WCDMA post-LNA filter. The obtained fourth-order ladder filter uses two symmetrical cells (see Fig. 8.5a). To have only two different BAW resonators for this four poles filter as used in a classical BAW filter and thus to simplify the integration process, the $-C_{23}$ capacitors have been artificially introduced in each half-arm of the input–output ladder filter. Then, in order to recover a new schematic equivalent with the original one, new elements (C_{ti} and R_{ti}) must be introduced in the new schematic. The new capacitors C_t and the new source and load resistances R_t are defined as follows (see Fig. 8.5b):

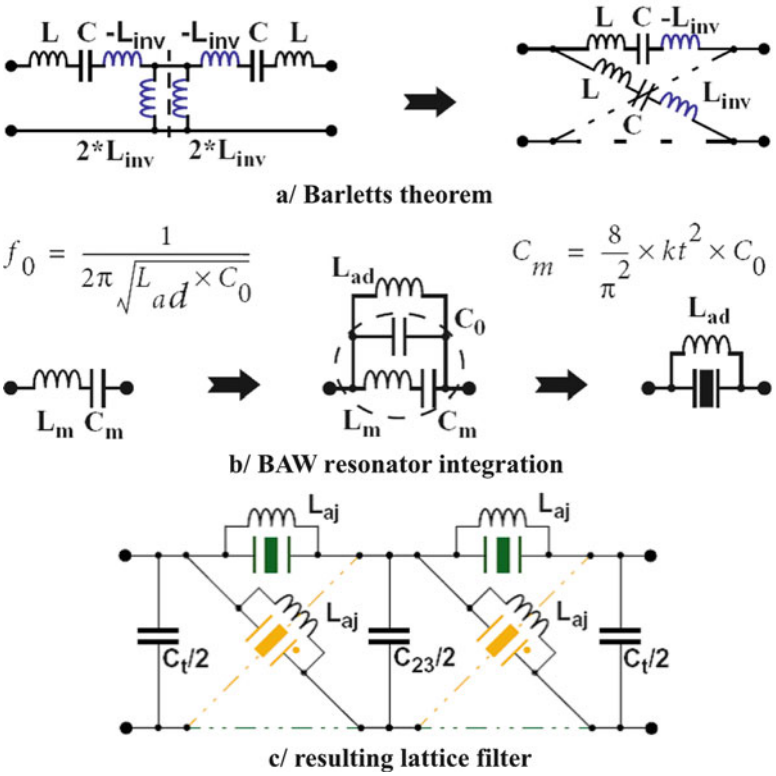

Fig. 8.6 Filter synthesis method. ©2008 IEEE. Reprinted, with permission, from [4]

$$C_{ti} = \frac{C_{23}}{1 + (\omega_0 C_{23} R_i)^2}, \quad i = S, L, \tag{8.4}$$

$$R_{ti} = R_i + \frac{1}{R_i(\omega_0 C_{23})^2}, \tag{8.5}$$

where R_i is either the source or the load resistance.

As a differential structure is less sensitive to substrate noise, ensures high second-order linearity, and eliminates the needs of a single to differential conversion, "Bartlett's Bisection Theorem" [2] (see Fig. 8.6a) is then applied to transform the single ladder filter into a differential lattice one. By adding a shunt capacitance C_0 (Fig. 8.6b) to the classical LC resonant circuit, it is possible to generate a BAW resonator equivalent model. However, this modifies the filter response. That is why a shunt inductance L_{ad} is added to the schematic in order to resonate with C_0 allowing to obtain an equivalent circuit to the initial series LC resonator around the filter center frequency. Figure 8.6c presents the synthesized BAW filter [3].

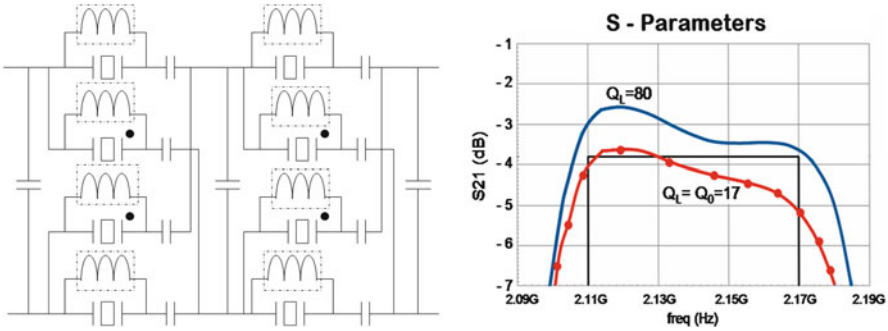

Fig. 8.7 (**a**) Electrically tunable filter. (**b**) Impact of the inductor's Q factor on in-band losses. ©2006 IEEE. Reprinted, with permission, from [5]

For a perfect equivalence, the inductance L_{ad} should be replaced by a negative capacitor in order to compensate C_0 and to be directly similar to a LC circuit without any disturbances. The final electrically tunable filter schematic is given in Fig. 8.7a. Figure 8.7b illustrates the need of high-Q inductors for an implementation with reduced in-band losses.

8.1.4 BAW Resonator Tuning

The goal is to correct a filter's process and temperature dispersions in the frame of the given filter mask. In classical use of ladder and lattice filters, both series and parallel frequencies of BAW resonator have to be shifted by the same ratio in order to shift the filter shape properly without any changes. A series capacitor will increase the series resonant frequency in the theoretical limit of the antiresonance frequency that remains constant (see Fig. 8.8a). A parallel one will decrease the antiresonant frequency while it makes the series frequency unchanged (see Fig. 8.8b).

A tuning component with an opposite phase allows to reach the opposite phenomena. A series inductor reduces the series frequency (see Fig. 8.8c), whereas a parallel one will increase the parallel frequency (see Fig. 8.8d). Nevertheless, inductors create additional resonances by interacting with the C_o capacitance of BAW resonator [6].

Integrated variable capacitors (varactors) in advanced process have only a limited tuning range. Controlling both resonant frequency f_s and antiresonant frequency f_p with only variable capacitances is not achievable on a large band.

Nevertheless, as seen before, the second principle of lattice filter appears as a more interesting alternative. Indeed, the use of the BAW resonator resonant frequency (f_s) by eliminating the antiresonant frequency offers a better opportunity to make lattice filter tunable. For this purpose, a parallel inductance will be used in order to resonate with C_o at f_p. Thus, only the resonant frequency (i.e., f_s)

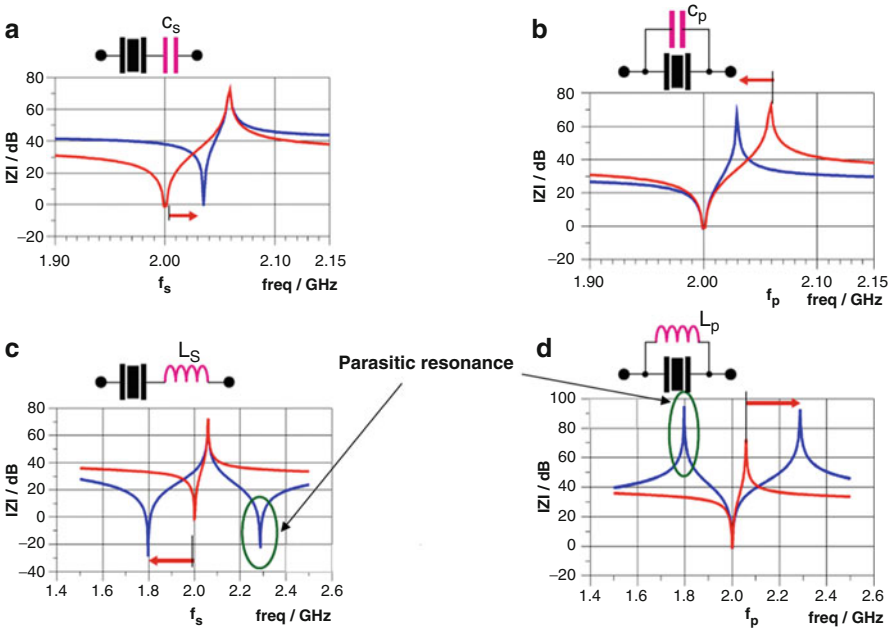

Fig. 8.8 Tuning a BAW resonator (**a**) with a series capacitance, (**b**) with a parallel capacitance, (**c**) with a series inductor, and (**d**) with a parallel inductor

will be exploited. Using this inductance pushes away f_p rather than eliminates it which is nevertheless sufficient. It then acts like a resonant cell whose resonant frequency will be adjusted by a series capacitance. Notice that switched capacitors could be a very good implementation for reconfigurable filters. However, care has to be taken about their Q factor. The next section will present the different solutions and considerations for the parallel component of the tuning cell.

8.1.5 Implementation of the Tuning Cell: Design of the Parallel Component

Table 8.1 summarizes the major trade-offs for the silicon integration of a high-Q inductance circuitry. A Si integrated spiral in the BEOL classically meets no more than 20 as a Q factor. An active inductor implemented using the gyrator technique [7] showed high power consumption and noise and linearity issues for the given application. The final choice went toward a spiral inductor with Q-enhancement circuitry.

Table 8.1 Trade-offs for the integration of a high-Q inductor

	Integrated spiral inductors	Active inductance	Q-enhanced inductors
Pros	Easy to implement	Low area	High Q (tunable)
	No power consumption	High Q (tunable)	
Cons	Large area	Prohibitive power consumption	Large area
	Low Q factor	Noise and linearity performances	Power consumption

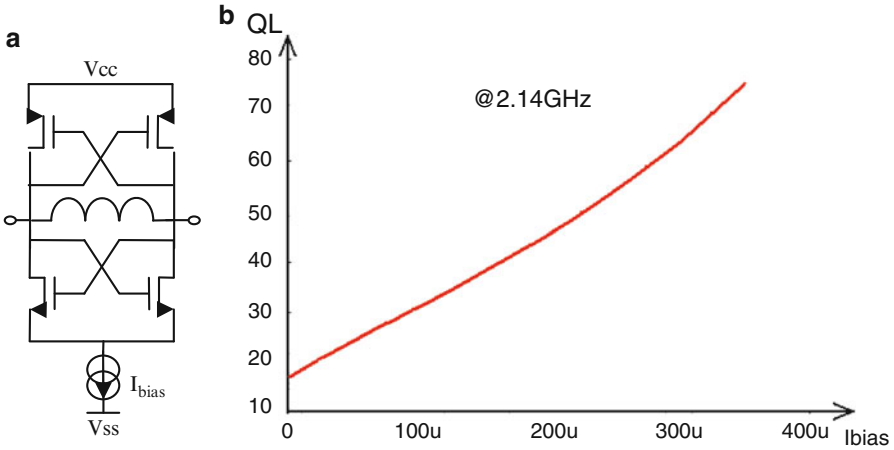

Fig. 8.9 (**a**) Q-enhanced inductor. (**b**) Variation of Q vs. I_{bias}

8.1.5.1 Q-Enhanced Inductor

Figure 8.9 presents a proposed circuit for a Q-enhanced inductor. Part of the losses of the spiral BEOL inductor is compensated by a current controlled negative resistor implemented using the gyrator technique.

This Q-enhancement circuitry allows to reach a Q factor of 80 for a current consumption of 350 µA. A negative resistance compensates the resistive part of the inductor in a narrow frequency band. A NMOS cross-coupled pair provides a $-2/g_m$ resistive impedance. Thus, adding a second PMOS cross-coupled pair allows increasing the negative resistance magnitude for the same tail current [8]. This biasing current controls the Q-factor inductance (see Fig. 8.9). In order to prevent the oscillations of shunt Q-enhanced inductor and the BAW resonator, the negative resistance magnitude must be less than the resistive part of the impedance at its characteristic frequencies.

In fact, a second resonance at $1/(2\pi\sqrt{L(C_{va}+C_0)})$ is created by adding the shunt inductance. Even though it does not impact on the filter response as this extra resonance is identical on parallel and series arms of the presented lattice filter, the antiresonant frequency can be eliminated by generating a negative capacitance $|-C| = C_0$.

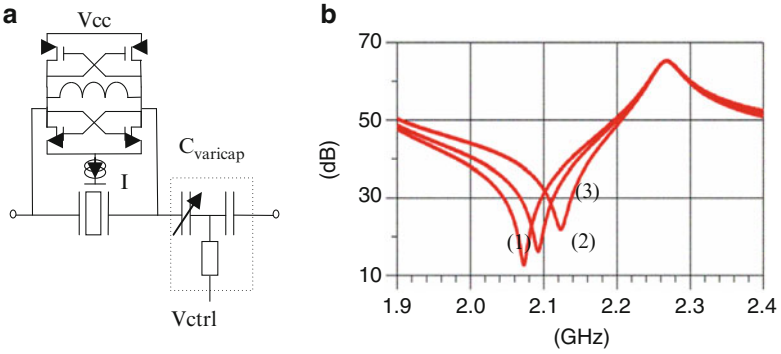

Fig. 8.10 Final tuning cell implementation: (**a**) scheme and (**b**) its impedance for different values of V_{ctrl}

8.1.6 Filter with Q-Enhanced Inductors

8.1.6.1 Physical Implementation of the Tuning Cell with Q-Enhanced Inductors

For a better distribution of parasitic capacitances, the varactor block has been placed after the couple {inductor, BAW resonator}. In fact, the inductance parasitic capacitances (around 300 fF) create a capacitive divider if the varactor is placed before this couple. In addition, the PMOS cross-coupled pair fixes the common mode that we exploit to bias the varactor. This permits to eliminate one of the coupling capacitances needed in series with the varactor. Thus, the varactor block consists of a metal-insulator-metal capacitor of 25 pF, a 100 kΩ poly-resistor and an NMOS varactor. Figure 8.10a shows the final tuning cell implementation, while Fig. 8.10b exhibits its simulated impedance for different varicap tuning voltage.

Each BAW resonator of the synthesized BAW filter is replaced by its respective tuning cell. Now, the filter tunability becomes a function of the tuning range of the varactor while the transfer function depends on only one control voltage.

8.1.6.2 Simulation Results

As expected, 50 Ωscattering parameter analysis exhibits good results on this filter. Indeed, in-band insertion losses are reduced by 1 dB thanks to the Q-enhanced inductance. The tuning cell permits to correct 1.4% of shift on the piezoelectric layer as Fig. 8.11 shows.

A third-order intermodulation of 16 dBm has been simulated in-band with two tones at 2.14 and 2.16 GHz. A 5.2-dB noise figure has also been simulated. Noise is mainly due to MOS transistors used in the negative resistance of the

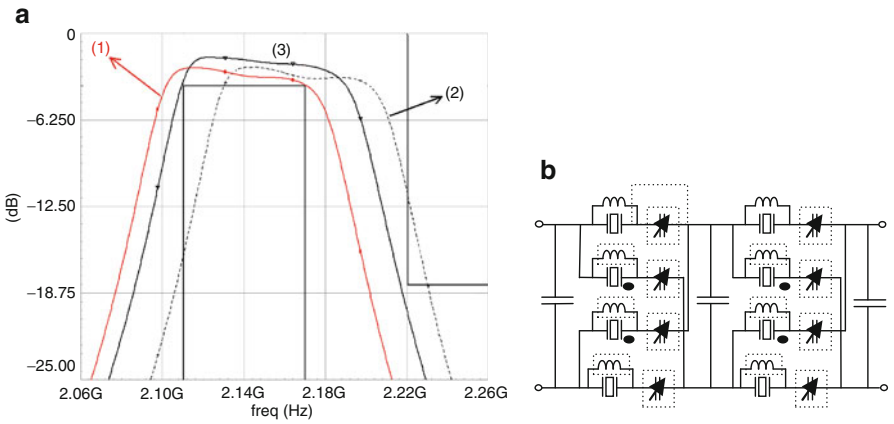

Fig. 8.11 (**a**) Tunable filter ©2006 IEEE. Reprinted, with permission, from [5]. (**b**) (1) Nominal filter response, (2) Filter response with BAW resonator presenting 1.4% shift on the piezoelectric layer thickness, (3) Filter response with BAW resonator presenting 1.4% shift on the piezoelectric layer thickness after correction by the varactor control voltage

Q-enhanced inductance. Finally, it is interesting to notice that the low passband cutoff frequency of the proposed tunable filter is defined at the frequency at which series and parallel impedances are equal in magnitude but opposite in phase.

To demonstrate the validity of this tunable filter, two chips have been implemented: the first one contains the 8 SMRs whereas the second one is the BiCMOS 0.25-μm SiGe technology part. They will be assembled by thermal compression thanks to a flip-chip bumping process. While targeted at above-IC BAW technology, we used flip-chip technology here to assemble an IC-die with an SMR die to demonstrate the validity of the concept (cf. Fig. 8.12). However, the flip-chip assembly has a negative impact on the insertion loss. Indeed, BUMP PADs add interconnection losses and do not permit to optimize filter size. Their design rules impose spatial constraints that allow balancing the pressure during the assembling part. Besides, this filter architecture is compounded of eight high-value inductances whose size is critical. And, to avoid the self-coupling, a minimum gap is effectively required hence contributing to the increase in the filter size.

8.1.6.3 Measurement Results

The electrical characteristics of the SMRs used in the filter are summarized in Table 8.2.

As stated in the first part of this chapter, the in-band ripple is dependent on the ratio of filter's branch impedances and also on the phase difference. The measured filter in-band ripple (1.5 dB) is larger than the simulated one (0.8 dB) and may be explained by the slight discrepancies between targeted and measured

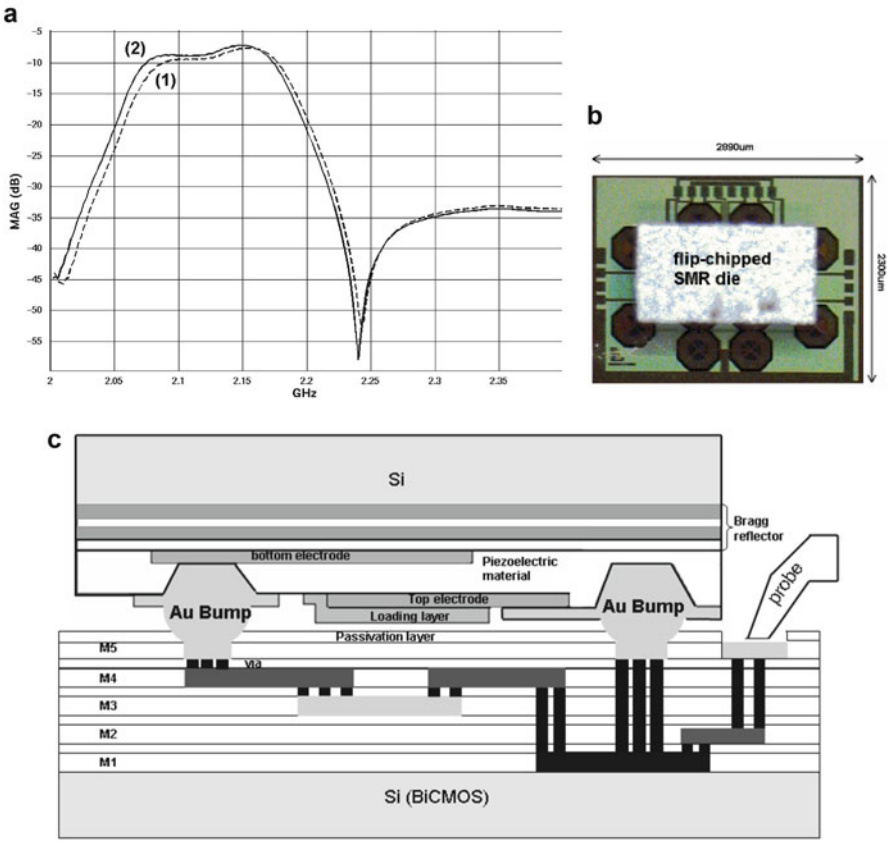

Fig. 8.12 (**a**) Measured maximum available gain for the RF filter transfer function (Curves (1) and (2)) corresponding to $V_{ctrl} = 0\,V$ and $V_{ctrl} = 2.5\,V$, respectively. (**b**) Photomicrograph of BAW/IC assembly ©2006 IEEE. Reprinted, with permission, from [5]. (**c**) Cross section of BAW/IC assembly with gold bump process ©2007 IEEE. Reprinted, with permission, from [9]

Table 8.2 SMR characterizations

SMR Resonators	C_o (pF)	Impedance at 2.4 GHz	Parallel SMR f_p (GHz)	Series SMR f_s (GHz)
Targeted	1.37	54	2.0528	2.1158
Measured	1.77	42	2.050	2.124

BAW resonator parameters in Table 8.2. The measured maximum available gain is reported in Fig. 8.12a. An out-of-band rejection of 28 dB has been measured over a wide frequency band. The first prototype exhibits an extra 2-dB insertion loss compared to the post layout simulations [5].

Several factors may be responsible for this in-band gain degradation. First of all, coupling between BiCMOS on-chip inductors is enhanced by the metallic plate of

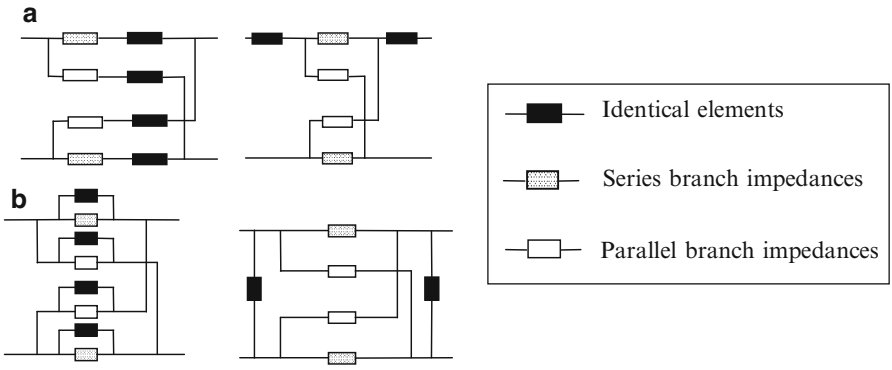

Fig. 8.13 Network theorem. (**a**) Removal of a series element. (**b**) Removal of a parallel element

the flip-chipped die. Secondly, the large chip area implies significant parasitics on the access lines. Finally, the bump access resistance degrades the BAW resonator's Q factor.

The filter's center frequency may be tuned over a 0.3% relative frequency band that corresponds to a correction of a 0.6% error on the piezoelectric layer. Moreover, two notch frequencies due to mismatch between impedances of series and parallel branches appear near the passband.

Further optimizations should considerably improve performances of such filters in the future. Extra efforts on the layout of a second tunable BAW filter have led to constant and lower insertion losses in the filter's bandwidth. Indeed, an improved filter structure using less inductors, an increase in the BAW resonator's Q factor and above-IC integration of such filters will also minimize losses due to the parasitic elements.

8.1.6.4 Filter Architecture with Reduced Inductor Count

The filter area can be reduced by simply decreasing the number of inductances. The application of two useful network theorems [2] allows the conversion of the 8-inductance filter into a 4-inductance filter. Indeed, identical series or parallel elements can be moved out of the series and parallel arms by strict equivalences (see Fig. 8.13). This allows reducing the filter area by decreasing the number of inductors.

Then, inductances have been halved by the parallel removal theorem. Nevertheless, in order to keep a symmetrical fully differential filter, the varactor blocks have been duplicated on the opposite series branch. A filter with two classical lattice cells (only composed of BAW resonators on its arms) linked by inductance-varactor blocks has been obtained.

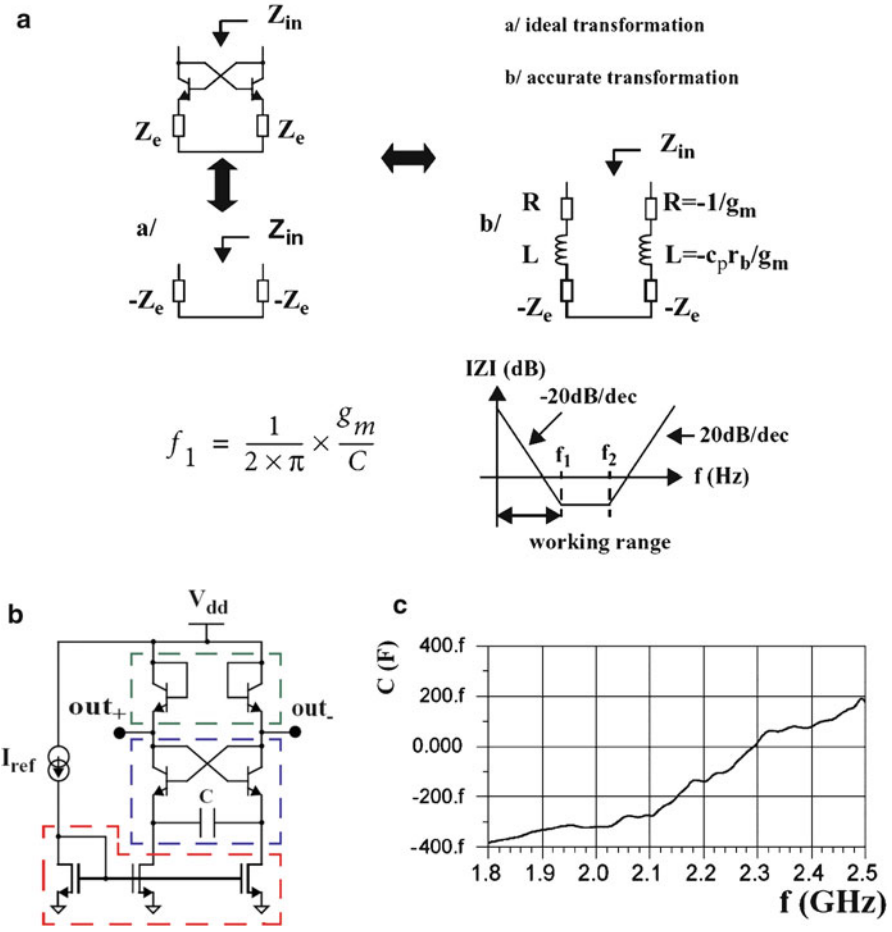

Fig. 8.14 Negative capacitance: (**a**) principle (©2008 IEEE. Reprinted, with permission, from [4]). (**b**) implemented circuit, and (**c**) its measurement result

In fact, lower insertion losses yield higher tunability. The resulting filter is tunable on a $\pm 1\%$ tuning range as illustrated in Fig. 8.12a; thus, the targeted process or temperature deviations can be totally compensated.

8.1.6.5 Inductorless Filter

As seen in Fig. 8.6, a differential active component with a 90° phase can replace the parallel spiral inductors as done in [4] without introducing extra parasitic resonances. This second type of tunable BAW resonator is presented in Fig. 8.14a. It shows the need of an active circuitry equivalent to a negative capacitor in a

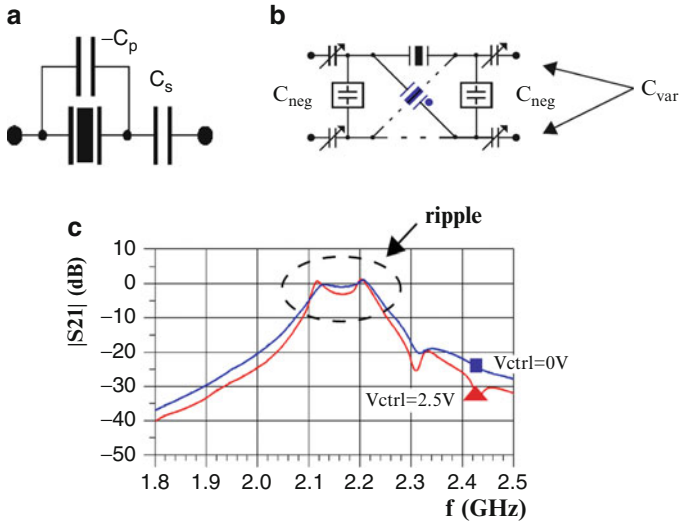

Fig. 8.15 (**a**) Tunable BAW resonator with negative capacitance. (**b**) Topology of the implemented filter. (**c**) Measurement results of a tunable BAW filter with negative capacitances (the curve indicated by the square corresponds to $V_{ctrl} = 0\,\text{V}$, whereas curve indicated by the triangle corresponds to $V_{ctrl} = 2.5\,\text{V}$). ©2008 IEEE. Reprinted, with permission, from [4]

given frequency band. For instance, a negative capacitance transformation may be obtained with a cross-coupled transistor pair (see Fig. 8.14). If $Z_e = 1/(j\omega C)$, ideally the input impedance looking into the cross-coupled pair will be equivalent to a negative purely capacitive impedance $-C/2$ (see Fig. 8.14a). It is convenient to take into account a more accurate transformation (see Fig. 8.14b) with g_m, c_p, and r_b (classical small-signal parameters of a bipolar transistor). The transformed negative impedance can now be expressed as the sum of a negative resistance, an inductive term and a negative capacitor.

A compromise needs to be found on the type of active device to be used for the impedance inverter. The maximum operating frequency f_1 implies a significantly high value of g_m to work at gigahertz frequencies. For reason of efficiency, bipolar transistors are preferred to MOS devices because of a larger g_m value for an equivalent current consumption. Moreover, the inductive effect implies the need for a small r_b value. In that case, a MOS transistor with an optimum layout provides a reduced r_g value instead of a bipolar ones.

This structure of negative capacitance presents a negative resistance too, thus permitting to enhance the equivalent Q factor of any element connected in parallel at the output nodes. The diode-mounted NPN bipolar transistors are exploited to obtain on one hand a potential drop and on the other hand a pole that increases the voltage operating range.

Figure 8.15 shows the implemented schematic for the negative capacitance circuit and the measured results. Post-measurement analysis has pointed out the

existence of several discrepancies implying a lower $|C|$ value than initially expected. Nevertheless, a slightly modified redesign may permit to obtain a value around -1 pF for a 200-$mV_{p\text{-}p}$ input dynamic range.

This second type of tunable BAW resonator (cf. Fig. 8.15a) has been then integrated into a one-section lattice filter, as depicted in Fig. 8.15b.

The level of insertion losses is only 1 dB that allows satisfying the WCDMA specifications, and the 15-MHz tuning range is sufficient to compensate the BAW process variations. This architecture presents an excellent out-of-band rejection with a minimum of -50 dB at 1 GHz from the central frequency (see Fig. 8.15c). This new 2-GHz bandpass filter presents a 1-mm^2 size area. It consumes 32 mW and an in-band third-order intermodulation of -9.5 dBm has been measured.

8.2 Tuning Circuitry

8.2.1 Preliminary Discussion

Finally, the direct tuning method (cf. Fig. 8.16a) implies that the electrical block to be tuned is taken off the signal path during the tuning period [10, 11]. This is incompatible with mobile communication standards with time division multiplexing mode. Thus, such a calibration becomes inherently impossible to conceive for numerous standards unless the calibration can be operated before starting all communication phases.

The master/slave technique corresponds to an indirect tuning method (cf. Fig. 8.16b). This one is very convenient to address the tuning of filters and is particularly used with G_m–C filters [12]. One of the approaches used for the master/slave technique is to lock a given device referred to as the master circuit with respect to a fixed time reference using well-known frequency synthesis methods like PLLs and to use the generated quantity (usually the control voltage of the VCO) to tune the slave circuit, which has to be composed of the same basic elements as the master circuit. Parasitic elements added by the tuning circuitry can make the

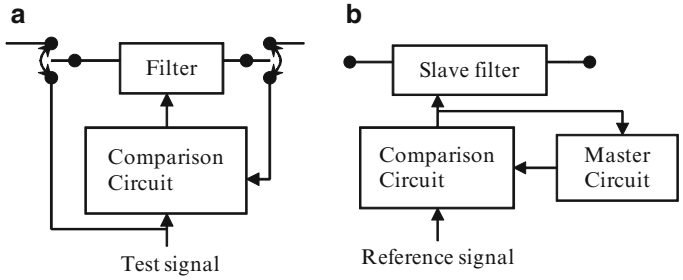

Fig. 8.16 (**a**) Direct tuning principle. (**b**) Indirect tuning principle

master cell environment varied with respect to the slave cell environment and thus generate a shift in the master cell impedance, thus providing a bad correction. The master circuit has to be representative of the slave circuit and to be as sensitive as slave cell in a same amount.

Several tuning strategies may be foreseen for the automatic tuning of the previously presented filters. The choice of the master and the slave circuit turns out to be crucial. Obviously, for reason of size, the filter cannot be duplicated. In this case, the different decision criteria can be the following:

- A bandpass filter has its central frequency defined when its phase is zero. A direct tuning method could be exploited to detect the phase difference between the output signal and the test signal feeding the filter's input.
- As a control voltage allows to shift the BAW resonator impedances by the same amount, one of the resonant frequencies of one of the tuning cells can be also tuned and controlled by an indirect tuning system.
- Finally, as the low passband cutoff frequency of the proposed tunable filter is defined at the frequency at which series and parallel impedances are equal in magnitude but opposite in phase, a third solution consists of detecting impedance magnitude by an indirect tuning method.

In fact, as the insertion losses of the filter arouse a shift between the frequency at which the filter phase is null and its central frequency, the first proposed solution turns out difficult to be implemented. Moreover, an extra, complex loop would be needed to correct such a phase shift. On the other hand, distortion harmonics in the reference signal will also create phase error by interfering with the phase detector [11]. Therefore, only the two other solutions will be discussed in the following section. They are applied to the BAW tunable filter given in Fig. 8.11.

8.2.2 Discussion on the Tuning Methods

8.2.2.1 Indirect Tuning Method: PLL with a VCO Master Cell

Among indirect tuning methods, the tuning of the resonant frequency of one of the BAW-based resonators (in series or in parallel branches) can be implemented. The BAW resonators can be exploited in a VCO structure and inserted in a PLL. The type of the VCOs (series or parallel tank) will define the resonant frequency to use. However, according to the operating principle of the filter, it turns out to be mandatory to use the resonant frequency. Just notice that the required accuracy on the phase detector is no more necessary in this PLL solution (see Fig. 8.17a).

The efficiency of this tuning technique implies several operation mode constraints. Firstly, oscillation amplitude has to be controlled to place the master and the slave circuit in same operating mode. The negative resistance of the Q-enhancement inductors can generate nonlinearities and disturb the lock of the PLL. Moreover, the VCO has to be responsive to the filter sensitivity to process and

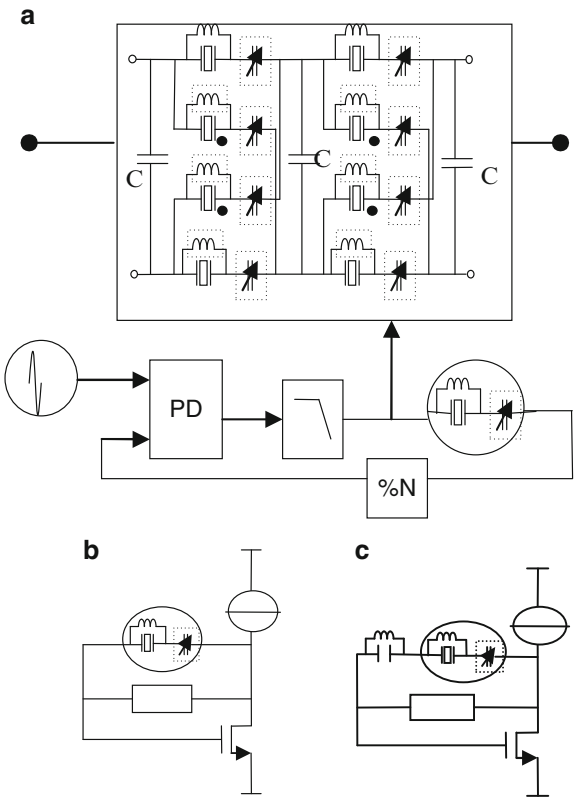

Fig. 8.17 Example of master/slave tuning circuit (**a**) with a matched Pierce VCO on the parasitic (**b**) and on the useful (**c**) series resonant frequency

temperature frequency deviations. The tuning cell exhibits two resonant frequencies. The parasitic one is determined by the inductor value and by its two capacitors C_0 and varactor capacitance value. This lower resonant frequency is very attractive for reaching lower power consumption. Moreover, the plate electrode capacitor C_0 is dependent on the thickness of BAW resonators and thus could well characterize the main BAW process dispersions. The main drawback is that the resonant frequency of BAW resonators is also defined by all the other stacked materials. Therefore, it seems to be unavoidable to exploit the useful resonant frequency. The series resonant frequency can be used by placing the tuning cell in the direct feedback of the VCO as can be done in Pierce configuration VCO (see Fig. 8.17b). Thus, naturally, the VCO will oscillate at the frequency requiring less energy (which is the parasitic frequency), and an extra trap resonant circuit has to be used to force oscillation at the second resonant frequency (see Fig. 8.17c). However, this extra circuit inserts parasitic capacitances contributing to supplementary mismatches and thus to a quasi systematic tuning error.

This master/slave technique cannot be applied with this tuning cell but can be convenient with the tuning cell using a negative capacitance. However, even in this case, oscillating at 2 GHz needs relatively high area transistors loading directly the

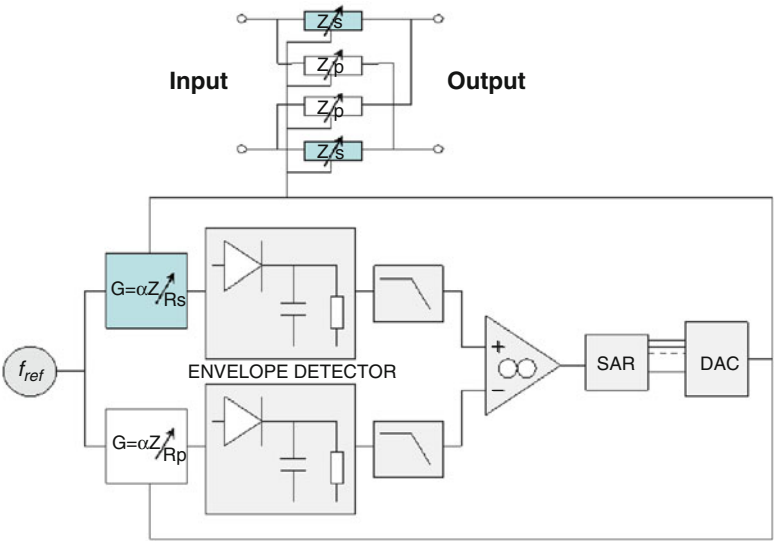

Fig. 8.18 Tuning circuitry principle

tuning cell and shifting the oscillation frequency toward the effective series resonant frequency of filter's BAW resonators. Matching the slave filter to the master VCO at gigahertz frequencies is difficult. Their respective structures are very different and do not naturally match. Extra capacitors in the oscillator core give rise to frequency pulling which results in a nonnegligible tuning error.

An alternative solution is possible with an envelope detection to determine the characteristic frequencies of the filter.

8.2.2.2 Indirect Tuning Method: Frequency-Locked Loop by Envelope Detection

For this tuning method, we have to detect the frequency at which the two resonant structures are equal in terms of impedance magnitude. If we limit the detection frequency range, we are located in the frequency interval where both impedances have opposite phase. In this way, by using two gain blocks proportional to Z_s (impedance of the series tuning cell) and respectively to Z_p (impedance of the parallel tuning cell) and by injecting a signal at a reference frequency, we are able to compare the level of both impedances at this frequency (see Fig. 8.18) [13]. Associated to a low-pass filter, an envelope detector will provide this information. A true bit (=1) will be generated by a comparator if the difference between Z_s and Z_p (noted Δ) is positive. Otherwise, a false bit (=0) will be generated.

The comparator output controls a successive approximation register (SAR) associated to a DAC, which will adjust the tuning voltage to the appropriate

value, i.e., where both impedances Z_s and Z_p are equal in magnitude. Indeed, the low-frequency clocked SAR will increment the central frequency's control voltage by addressing adequate bits to the DAC block following a dichotomy tuning law. The resonant structures will be shifted by the same amount while always being fed by the reference signal. This tuning method will be applied till the sign of Δ changes. The final control voltage value to be applied to the filter has been thus found. Then, by injecting this control voltage to the filter's control voltage, the filter's central frequency is maintained close to the reference frequency. Moreover, a supplementary bit (CLEAR) may be added in the SAR module to launch the tuning of the filter. According to the sensitivity of the filter, it is possible to tune it continuously and automatically or only punctually.

Furthermore, the accuracy of the tuning circuitry is highly linked to the matching of the gain blocks proportional to Z_s and Z_p and the resonant structures used effectively in the filter. It also depends on the reference clock. Indeed, because of time constant, the tuning circuitry needs to have enough time to reach the steady state and then to come to the good detection. The slower the clock, the more exact the decision. Finally, the DAC will define the steps of the control voltage and has to be adapted according to the required sensitivity.

8.2.3 Implementation of the Frequency-Locked Loop Using Envelope Detection

8.2.3.1 Principle of the Implemented FLL

The gain is proportional to the impedances of each branch as given in the following equation:

$$G_{s,p} = \frac{Z_{s,p}}{R + Z_{s,p}}, \tag{8.6}$$

where R is placed in the direct path and $Z_{s,p}$ is grounded (see Fig. 8.19).

To match the slave to the master cell, Z_p and Z_s have to be loaded by the same capacitors seen by the respective slave impedances employed in the filter to be tuned. On the other hand, in order to increase the accuracy of the detection, the resulting output signal will be amplified before sampling its envelope magnitude. Figure 8.19a exposes the first stage of the tuning circuitry with its amplifier stage associated to the envelope detector. It allows to reduce the constraints of high-frequency design by down-converting the tuning operating frequency in a MHz frequency range. This pseudo-differential part has to be well matched. Indeed, the error committed on the envelope detection is, in the first order, the same on the two paths, reducing the error on the decision. This part of the circuit consumes 2.54 mA

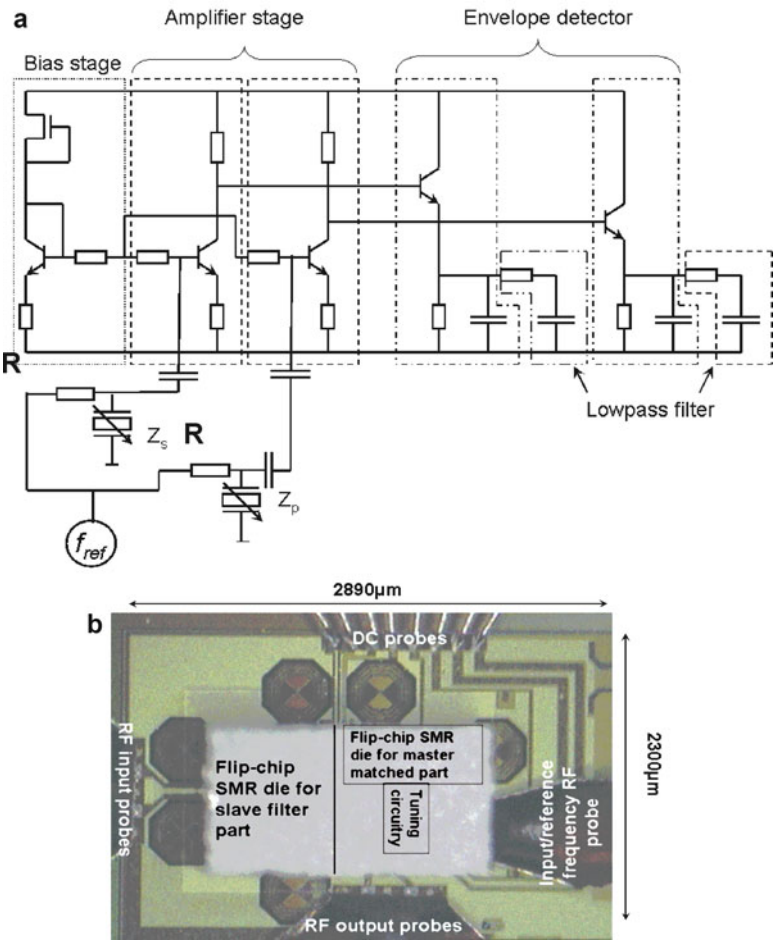

Fig. 8.19 (**a**) Down-converted part of the tuning circuitry and (**b**) microphotograph of the master/slave system. ©2007 IEEE. Reprinted, with permission, from [14]

current under 2.5 V and exhibits a 3-dB gain. The envelope detector is directly dc coupled with the amplifier output. It exploits the PN junction of a common collector bipolar transistor as a diode (see Fig. 8.19a).

An OTA consuming 220 µA under 2.5 V has been designed as a comparator and provide the decision bit equal to one when the Z_s magnitude is greater than that of Z_p, otherwise a bit equal to zero. This bit called RESULT will control the SAR whose time reference will clock the filter tuning. For the proposed SAR design, a single D-type flip-flop is used in each bit cell which functions both as sequencer and code register. This type of design is often referred to as the sequencer/code register design [15]. It consumes 40 µA and can be clocked from 1 to 5 MHz. Furthermore, to reduce the locking time at the beginning of the tuning sequence, the SAR is

initialized to the midvalue of the control voltage interval, i.e. MSB = 1, all other bits are 0. The bits are then adjusted according to the comparator output value starting from the MSB going toward the LSB, the final value of the tuning voltage being within 1 LSB from the ideal value [14].

The full circuit is assembled using a SiP method where the 0.25-µm BiCMOS die and the BAW SMR die are contacted by a flip-chip process (see Fig. 8.19b). The first chip contains 6 SMRs (four for the filter and two for the master cell) in an AlN process provided by CEA/LETI and STMicroelectronics, whereas the second one is processed in STMicroelectronics 0.25-µm SiGe:C BiCMOS technology.

8.2.3.2 Measurement Results

For testability reasons, the reference frequency has not been implemented and would be provided by an external source. This source has been swept from 2.07 to 2.09 GHz with a step of 100 kHz and a 1-$V_{p\text{-}p}$ amplitude. As depicted in Fig. 8.20a, the displayed results represent the behavior of the filter control voltage versus the reference frequency whereas the variation of the corresponding central frequency is also drawn. The measured filter presents a constant 104-MHz bandwidth, and its central frequency is controlled on a 10.5-MHz frequency range. Finally, the obtained error is about 3.6 MHz that is less than 0.2% error whereas the tuning step is 100 kHz. Furthermore, the tunability is facing the nonlinear capacitance variation of varactors toward its control voltage. Implementing a nonlinear step for the DAC could sort out such an issue.

Two reference clocks (1 and 5 MHz) have been tested. A quasi systematic error is committed by the use of a faster clock because the decision has been done before achieving the steady state. Figure 8.20b capture has been done for a 2.083-GHz input frequency, and the measured signal is the filter control voltage.

One can observe the dichotomy function of the SAR-DAC stage. Rise time is defined by the product of the decoupling capacitor's varactor to load and its decoupling resistors. It actually determines the clock frequency to use. Nevertheless, the establishing time of the tuning circuitry is function of the error between the targeted central frequency and the effective one. This value is included between 2 and 4 µs. Finally, according to the applied correction voltage, the tuning circuitry consumes from 2.54 to 4 mA. However, if a punctual correction is sufficient, the tuning circuitry can be turned off after storing the control voltage.

8.3 Conclusion and Perspectives

Both tuning cell prototypes—with Q-enhanced inductors and with negative capacitors—validate the feasibility of synthesizing and designing tunable BAW bandpass filter which is able to compensate the AlN process and temperature dispersions. Measurements have showed good agreement with post-layout

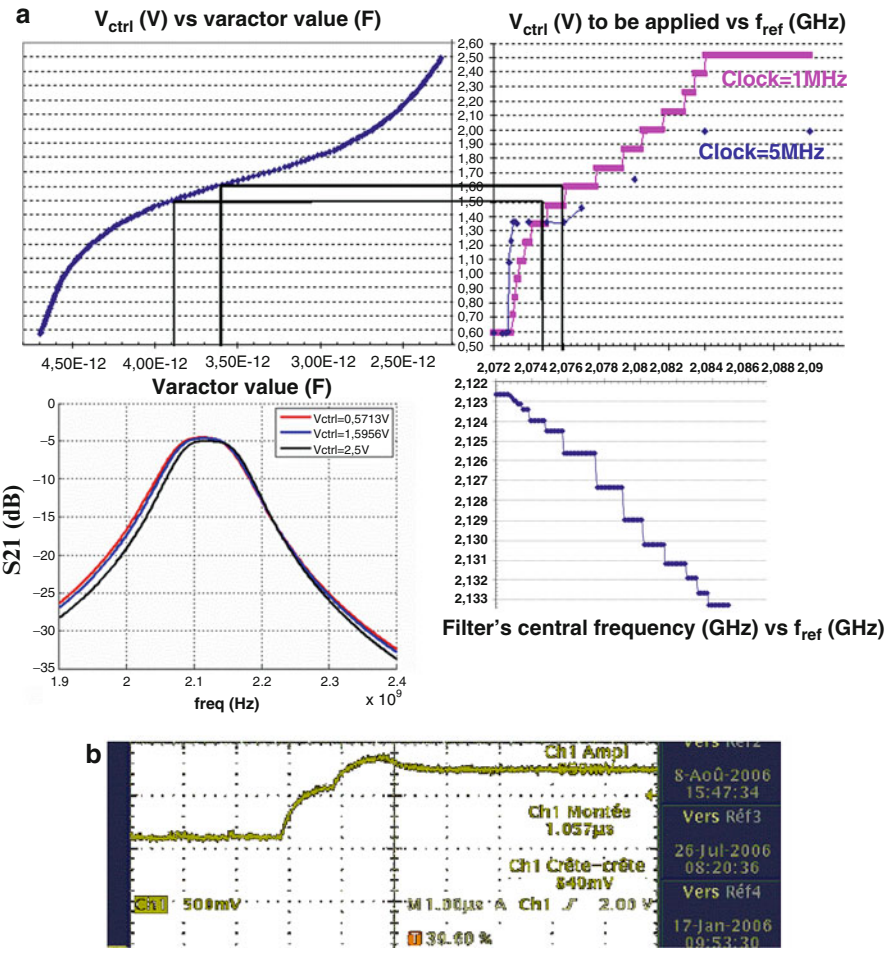

Fig. 8.20 Measurement results

simulations. But further work is still needed to decrease the in-band insertion losses by the use of passive high-Q components. Moreover, it offers new perspectives on filtering issues; for example, in the case of the tuning cell using a shunt inductor, by taking a different value for inductances or varactors from the series resonator to the parallel one, an extra resonant frequency can be exploited in order to obtain a second bandpass filter at lower frequencies [16]. However, it still requires high-Q parasitic resonant frequencies and thus the use of high-Q components in a very large frequency band. In order to automatically correct the frequency deviations, we need to associate a tuning circuitry to this BAW tunable filter. Therefore, a master/slave technique which allows to digitally tune a tunable BAW filter's central frequency with respect to a single available reference clock has been proposed.

This system is aimed at correcting process and temperature dispersions and at tuning all types of lattice filters accurately. Instead of relying on, the sometimes complex, PLL tunability, the whole flexibility has been moved to the proposed master/slave scheme, which consists of envelope detection. The control frequency range is limited by the tuning varactor range and can be made larger by the use of a matrix of switchable capacitors. It could be suitable for a reconfigurable application if a convenient filter is designed.

Thus, the intimate BAW/IC cointegration has been demonstrated with a STMicroelectronics 0.25-µm 2.5 V CMOS process and SMR resonators. From now on, individual BAW devices may be seen as intrinsic parts of an IC process just like an inductor or a capacitor. Thus, we may think of redefined circuit architectures taking benefit of the large Q factor of such resonators. Typical examples of system blocks where BAW devices are attractive are oscillators [9] and filters.

References

1. A.I. Zverev, *Handbook of Filter Synthesis* (Wiley, New York, 1967)
2. R.G. Kinsman, *Crystal Filters: Design, Manufacturing and Application* (Wiley, New York, 1987)
3. C. Tilhac, et al., in *IEEJ Analog VLSI Workshop*, 2005
4. C. Tilhac, et al., in *Proceedings Radio Frequency Integrated Circuits Symposium (RFIC)*, 2008
5. S. Razafimandimby, in *Proceedings European Solid-State Circuits Conf. (ESSCIRC)*, Montreux, 2006
6. S. Razafimandimby, et al., *Electronic Circuit Comprising an Adjustable Resonator*, 2007
7. B. Nauta, IEEE J. Solid State Circ. **27**(2) (1992)
8. S. Razafimandimby, et al., A 2GHz 0.25µm SiGe BiCMOS Oscillator with Flip-Chip Mounted BAW Resonator in *IEEE International Solid-State Circuit Conference (ISSCC) Digest of Technology papers*, San Francisco, 2007
9. S. Razafimandimby, et al., in *International Solid-State Circuit Conference (ISSCC) Digest of Technology Papers*, San Francisco, 2007
10. K.A. Kozma, D.A. Johns, A.S. Sedra, in *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 2, 1995, pp. 1037–1040
11. F. Krummenacher, N. Joehl, IEEE J. Solid State Circ. **23**(3), 750 (1988)
12. Y.P. Tsividis, IEEE J. Solid State Circ. **29**(3), 166 (1994)
13. S. Razafimandimby, et al., *Frequency Tuning Circuit for Lattice Filter*, US Patent publication number US2009072925, march 2009
14. S. Razafimandimby, et al., in *Proceedings European Solid-State Circuits Conference (ESSCIRC)*, Munich, 2007
15. T. Anderson, Comput. Des., pp. 81–86 (1972)
16. S. Razafimandimby, et al., *Filtering Circuit Fitted with Acoustic Resonators*, US Patent number US20080024244, january 2008

Part III
MEMS-Based Systems

Chapter 9
A MEMS-Enabled Two-Receiver Chipset for Asynchronous Networks

Brian P. Otis, Nathan Pletcher, and Jan Rabaey

Abstract Although technology scaling of CMOS devices has allowed great advances in device speed and digital computational efficiency, the quality and size of on-chip passive structures have not enjoyed similar scaling. As a result, the power consumption, physical footprint, and performance of modern integrated transceivers are typically limited by their passive components. In this chapter, we explore the benefits of augmenting CMOS ICs with high-quality RF MEMS resonant structures. In addition to improving the performance of standard RF building blocks, these RF MEMS components enable fundamentally new radio architectures. Various techniques for utilizing RF MEMS resonators will be discussed. We introduce two new receiver topologies based on these techniques: one is an always-on wake-up receiver with an uncertain intermediate frequency (IF); the other is a high-sensitivity super-regenerative receiver.

9.1 Introduction

Miniaturized, low-power wireless communication remains one of the great challenges of an increasingly connected world. Emerging applications in implantable human health monitoring, for example, require small form factor and infrequent battery replacement. Very aggressive applications may even necessitate a completely

B.P. Otis (✉)
University of Washington, Seattle, WA, USA
e-mail: botis@uw.edu

N. Pletcher
Qualcomm Incorporated, San Diego, CA, USA

J. Rabaey
University of California, Berkeley, CA, USA

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_9,

thin-film transceiver powered through solar or wireless energy sources. MEMS-based transceivers provide the potential to tackle both the size and power challenges presented by these new applications.

Section 9.2 discusses multiple techniques for using BAW resonators in RF building blocks, while Sect. 9.3 introduces two transceiver architectures that rely on these techniques to achieve low levels of power dissipation.

These two receivers illustrate two important capabilities of RF MEMS resonators. An uncertain-IF receiver leverages the extremely high-Q bandpass nature of the resonator to perform pre-filtering, relaxing the linearity requirements of the receiver front-end, and performing image rejection. A super-regenerative receiver, on the other hand, exploits the resonator response to stabilize a low-power, low-noise RF oscillator that forms the basis for achieving super-regenerative gain. In both cases, the resonator-enabled receiver achieves simplicity and efficiency that would be impossible using solely on-chip passive components. The theory, design, and results from both receivers will be presented. Finally, the system-level advantages of using an asynchronous two-receiver platform in an energy-constrained environment will be discussed.

9.2 Utilization of RF MEMS in Low-Power Transceivers

RF MEMS resonators were originally designed for applications like steep bandpass filtering or RF duplexers. However, they can also perform a variety of useful functions passively that are traditionally performed with power-hungry active circuitry. Three examples of how RF MEMS can be used are given here, along with design example case studies.

In recent years, RF MEMS research has yielded several different types of resonators useful as high-quality factor resonant passives. Attainable quality factors range from about 1,000 to more than 100,000. Most importantly, thin-film integrated circuit fabrication methods are used, offering low manufacturing cost and possible integration with CMOS. One of the most mature technologies, and the one chosen for this research, is known as bulk acoustic wave (BAW) resonators. A BAW resonator is built using a thin layer of piezoelectric material, usually aluminum nitride sandwiched between two metal electrodes. An electric signal on the electrodes excites a resonance in the structure whose frequency and quality factor depend on the thickness and geometry of the AlN layer. The resonator can be modeled quite accurately with the modified Butterworth-Van Dyke (mBVD) model comprising a series LCR circuit with lossy shunt capacitance, as shown in Fig. 9.1 [1].

Quality factors greater than 1,000 are readily achieved at frequencies in the gigahertz range. Measured data from a 2.4-GHz BAW resonator fabricated by Avago Technologies is shown in Fig. 9.2.

The dashed line is the measured impedance magnitude of a resonator with an active area of 10,000 μm^2. The solid line is measured from a resonator with twice the area (20,000 μm^2). Since the designer has the freedom to change the resonator layout

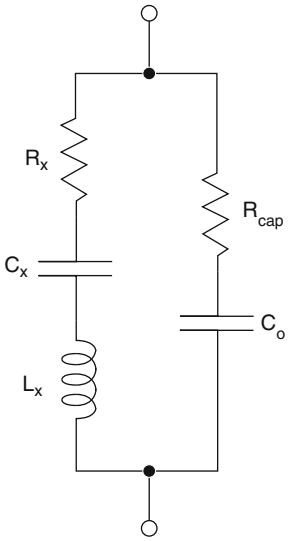

Fig. 9.1 mBVD circuit model of BAW resonator

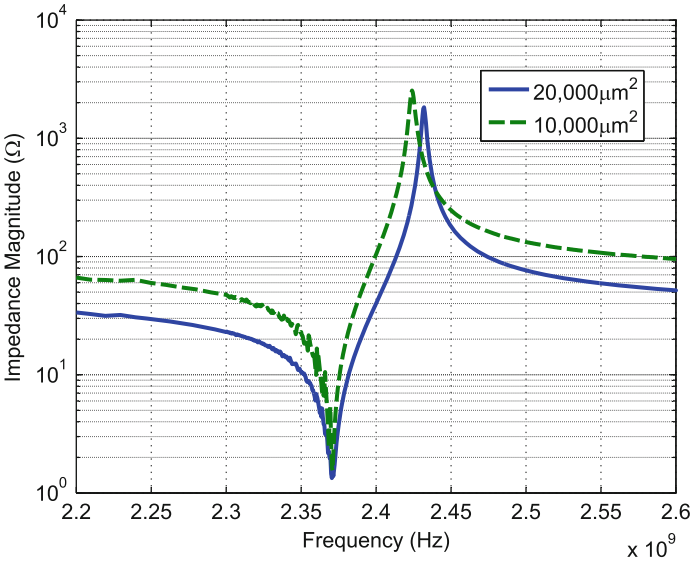

Fig. 9.2 Measured impedance magnitude for BAW resonators of two different areas

and area, it is useful to compare impedances as the area is varied. Off resonance, the impedance is dominated by the lossy capacitor C_o. As expected, the impedance of the large resonator is half that of the smaller device and the motional resistance R_x is slightly lower. Significant differences are seen at the parallel resonance. The parallel resistance of the smaller device is slightly higher but the larger device shows a higher

parallel resonance frequency f_p. This indicates that the larger device has a higher piezoelectric coupling coefficient k_t^2 and will exhibit better tolerance to additional capacitive loading due to the circuitry. It is clear that co-optimization between the IC and the resonator is desirable.

The high-Q nature of BAW resonators, even at high frequencies, allows the resonators to be incorporated into RF transceivers as either an oscillator frequency reference or as a filtering element. BAW devices are fabricated using entirely thin-film batch processing, enabling low cost and small size compared to traditional passives. In this transceiver design, the use of MEMS devices enables a simple, low-power implementation of both the receiver and transmitter in a small implementation volume. Two BAW resonators are sufficient to replace bulky quartz frequency references and no other off-chip passives are required.

9.2.1 Stabilization of Low-Power, Low-Noise RF Oscillators

Although BAW technologies are manufactured for bandpass RF ladder filter applications, they are especially useful in the design of low-power, low-noise GHz-range oscillators. The small size, high impedance at parallel resonance, and high-quality factor lead to extremely high-power/noise figures of merit. Frequency synthesis remains a critical problem in RF communications, especially in the low-power space. A few important considerations are:

1. Fabrication tolerance in parts per million (ppm). This metric predicts the stability of the oscillation frequency over process variations of all oscillator components (e.g., capacitors, resonators, transconductors).
2. Phase noise. The phase-noise performance of a sinusoid generator is measured in dB below the carrier per Hertz (dBc/Hz). Phase noise can be improved by increasing the resonator Q, increasing the oscillator signal power, reducing the oscillator noise factor, or enclosing the oscillator in a phase-locked loop referenced to an even-cleaner frequency source.
3. Long-term stability. The long-term stability of an oscillator is determined by the sensitivity of the oscillation frequency to temperature, supply voltage, and aging.
4. Power consumption. The power consumption of the frequency generator is of utmost importance in most applications. In the field of ubiquitous transceivers, extremely low-power dissipation is crucial for the success of the system.
5. Tuning range. Intentional frequency deviation is often desirable. For example, frequency calibration, channel switching, and modulation all require active control over the output frequency. In these systems, frequency control accuracy, resolution, settling time, and tuning range are important parameters.
6. Level of integration. Increasing the level of integration decreases the form factor of the system and decreases cost.

In [2], the first sub-mW BAW-based oscillator was presented. Figure 9.3 shows the schematic. It was fabricated in a 0.18-μm process and operated from

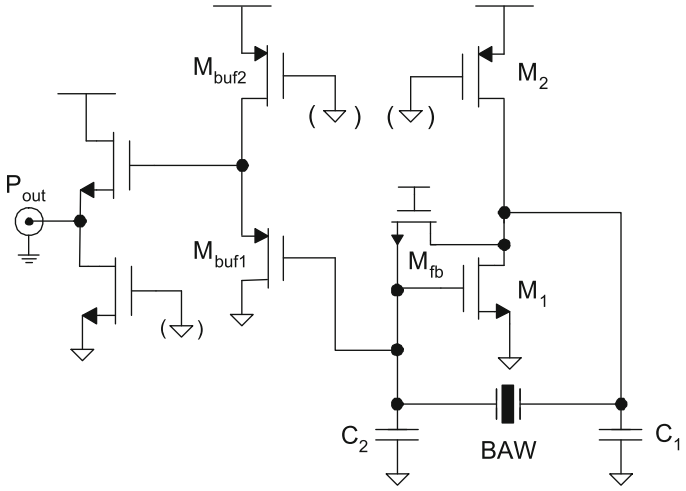

Fig. 9.3 Schematic of low-power BAW-based oscillator. ©2003 IEEE. Reprinted, with permission, from [2]

a 1-V supply. The oscillator consumed 300 μW and allowed a phase noise of −140 dBc/Hz at 1 MHz. Although a circuit designer using BAW resonators does not have the luxury of modifying the film thickness, there is an additional degree of freedom that affords the designer greater leverage in achieving an optimal design. For a given resonant frequency, it is possible to vary the active area of the resonator membrane. As the area of the resonator increases, the motional resistance (R_x in Fig. 9.1) decreases. Increasing the membrane area of the resonator increases the area of the parallel plate structure, thus increasing the feedthrough capacitance C_o. This, in turn, changes the oscillator tuning range and loaded parallel resistance. As discussed in the prior section, the area of the resonator should be optimally sized for a given circuit. For this oscillator, a 10,000-μm^2 resonator area yielded the optimal R_p, resulting in a minimization of power dissipation for the Pierce topology.

This oscillator prototype illustrates the advantages and disadvantages of free-running BAW frequency sources:
Advantages:

1. Spurious tones (common in frequency synthesizers) are absent.
2. The high-Q resonator ensures excellent open-loop phase noise.
3. Reduced power consumption and complexity over a frequency synthesizer.
4. Supply pushing and load pulling are greatly reduced due to the high resonator Q.

Disadvantages:

1. The manufacturing tolerance of BAW resonators can be improved by trimming but typically at the expense of fabrication cost.
2. The temperature coefficient of an uncompensated BAW resonator is approximately −30ppm/°C: much worse than a quartz resonator.

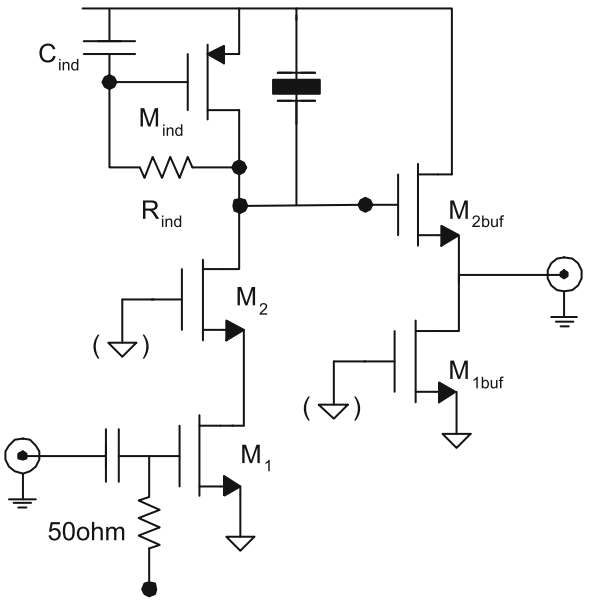

Fig. 9.4 Tuned RF amplifier schematic

3. The resonant frequency is often defined by film thickness, reducing the number of frequencies that can be fabricated on a single die.
4. Very poor tuning range (typically less than 1%) if capacitive tuning is used.

Much of the work that is currently underway in BAW-based transceiver design is centered around leveraging these benefits while mitigating these disadvantages. This oscillator forms the foundation of the super-regenerative receiver described in Sect. 9.3.2.

9.2.2 Tuning of High-Q RF Amplifiers and Design Example

High-Q resonators can be used as narrowband loads to perform filtering within RF amplifiers. The BAW resonator, used in the parallel resonance mode, performs high Q filtering by presenting a high impedance for a very narrow bandwidth about its parallel resonant frequency. However, for off-resonance frequencies, the resonator presents a $1/(j\omega C)$ impedance. Thus, at low frequencies (LF), the resonator presents a high impedance to the active circuitry. To provide bias current to the active devices and reduce LF gain, a low load impedance at LF is desirable. However, at the signal frequency, the impedance must be high to avoid de-tuning the BAW resonator. Thus, an inductive structure is a natural candidate. To avoid substantial de-tuning of the BAW resonator at 1.9 GHz, the bias device must present a high real impedance value (greater than 1,500 Ω) or so. An example of a BAW-tuned RF amplifier is shown in Fig. 9.4.

The RF amplifier consists of an NMOS cascode transconductance stage and a tuned load. A cascode structure is used to increase reverse isolation, improving amplifier stability. The tuned load consists of a BAW resonator and a self-biased PMOS-R-C that synthesizes a large (8.5 µH) inductance. This active inductor structure stabilizes the low-frequency bias point of the amplifier. The loaded quality factor of the tuned load is approximately 600, yielding an RF bandwidth of 3 MHz.

The use of an active inductor allows the realization of very high inductance values and easy control over the inductor Q. For this application, a large, low Q inductance is needed to provide high RF impedance while reducing the Q of the low-frequency parasitic resonance (occurring at approximately 30 MHz).

A stand-alone RF amplifier was designed and implemented to verify the performance of this structure. Transistor M_1 is operated in weak inversion for two main reasons: first, the degradation in f_T is not crucial because the gate capacitance of this amplifier is likely absorbed into the load of a prior stage. Secondly, the voltage gain of the stage is proportional to the transconductance of M_1, and g_m/I_d is maximized in lower levels of inversion. In this implementation, a g_m/I_d of approximately 20 is achieved with a bias current of 500 µA. The gate of cascode device M_2 is self-biased, and the layout of M_1 and M_2 is optimized for low-drain and interconnect parasitic capacitance.

A prototype was fabricated in a 0.13-µm CMOS process and assembled with a 1.9-GHz resonator. Figure 9.5 shows the measured s-parameters of the amplifier. The prototype used an on-chip 50-Ω input-match resistor, and as expected, the $|S_{11}|$ measurement shows a good wideband input match. It is interesting to note that the series and parallel resonances of the BAW resonator are visible in the $|S_{11}|$ plot. The forward gain, $|S_{21}|$, shows a peak gain of 10 dB (16 dB after de-embedding the output buffer). This gain matches our expectations given the following: $I_{bias} = 500\,\mu A$, $g_m/I_d = 20$, and the loaded resonator $R_p = 700\,\Omega$. At the series resonance, the resonator impedance drops to approximately $3\,\Omega$, where we would predict a rejection of $20 \cdot log(700/3) = 47$ dB compared to the parallel resonance. Our measurements reveal around 35 dB of rejection, likely due to parasitic inductance in the CMOS/BAW wirebonding. The $|S_{12}|$ and $|S_{22}|$ measurements show good reverse isolation and output matching, respectively.

9.2.3 Resonator Input Matching

The versatile BAW resonator is also easily embedded with input matching networks as a filtering element. As shown at the beginning of Sect. 9.2, the resonator exhibits as series resonance where the real impedance is low and a parallel resonance where it is high. In principle, the resonator can be used in either mode as a filtering element. In the series mode, the impedance is very low (about 1–4 Ω) and passes the signal. However, an additional impedance matching network must still be included to match subsequent circuit input impedance to the source. Using the resonator in the parallel

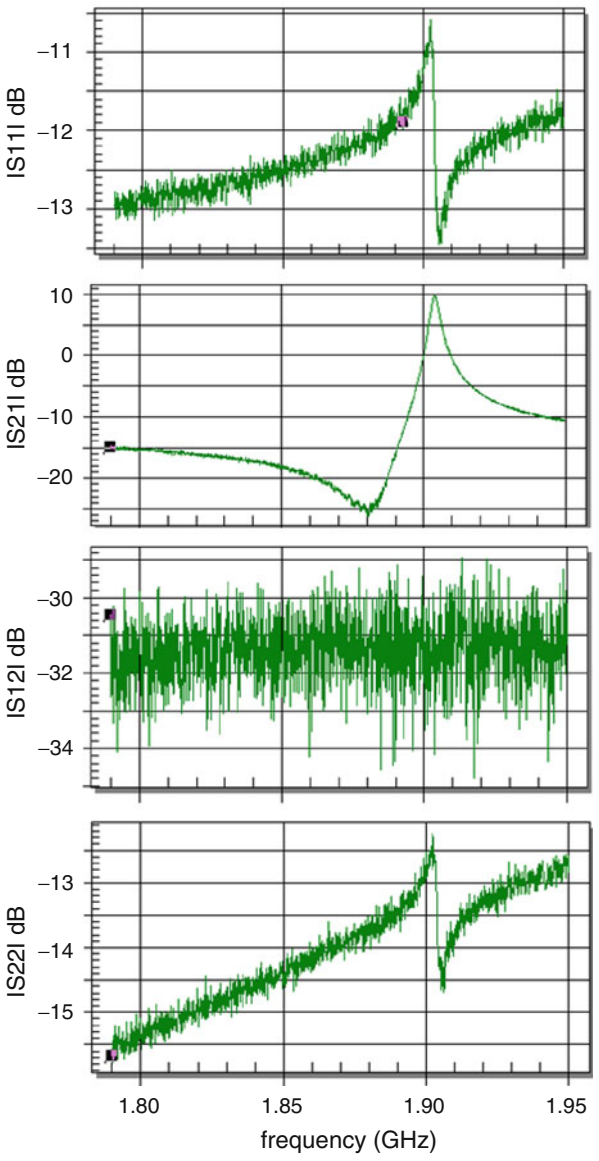

Fig. 9.5 Measured s-parameters of the RF amplifier

resonance mode, on the other hand, allows the resonator to become a part of the matching network. This can eliminate the need for large and lossy inductors in the matching network. The following example illustrates one possible configuration.

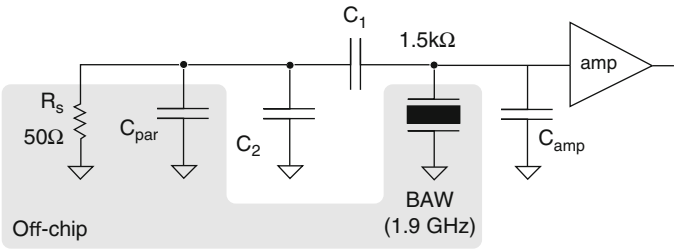

Fig. 9.6 Matching network with embedded BAW resonator

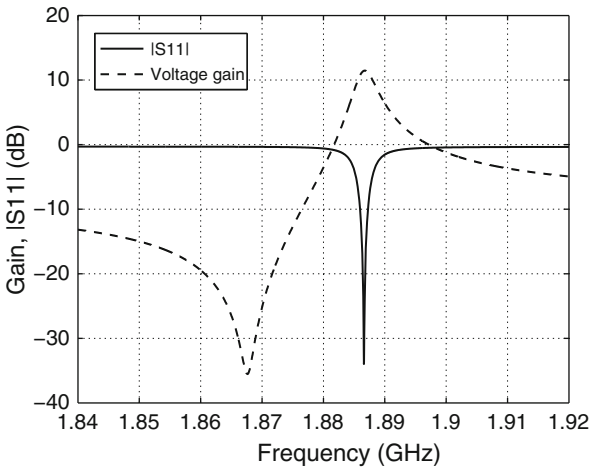

Fig. 9.7 Simulated return loss and voltage gain for BAW-based matching network

9.2.3.1 Passive Receiver Front-End Matching

Figure 9.6 shows a matching network using the BAW in shunt, parallel resonance mode where the impedance is high (about 1.5 kΩ).

A capacitive transformer (C_1 and C_2) transforms the low 50 Ω source impedance up to match the high impedance of the resonator. Any parasitic capacitance at the RF input port may be lumped into C_2, while the input capacitance of the subsequent circuit stage (an amplifier in this case) is conveniently absorbed by the large parallel capacitance of the BAW. This is an advantage for CMOS circuits, which naturally present a capacitive input impedance. An additional benefit is voltage gain through the network due to impedance transformation. Again, this is an advantage for properly designed CMOS circuits because the gates are sensitive to voltage. The passive voltage gain realized through the network is particularly valuable for ultra low-power receivers, where gain is costly from a power perspective. A simulation of the matching network voltage gain and return loss is shown in Fig. 9.7.

The network exhibits a good impedance match to the 50-Ω source and a passband bandwidth on the order of a few MHz.

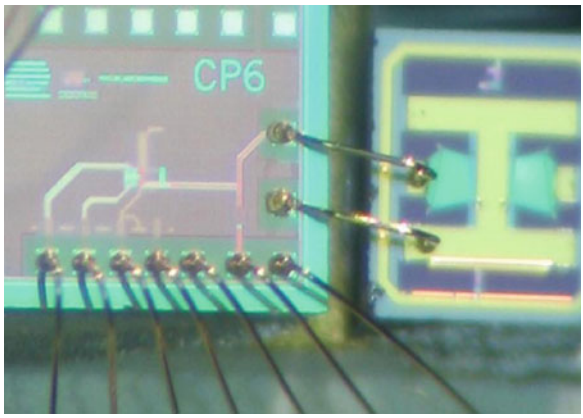

Fig. 9.8 Die photograph of passive receiver and BAW resonator

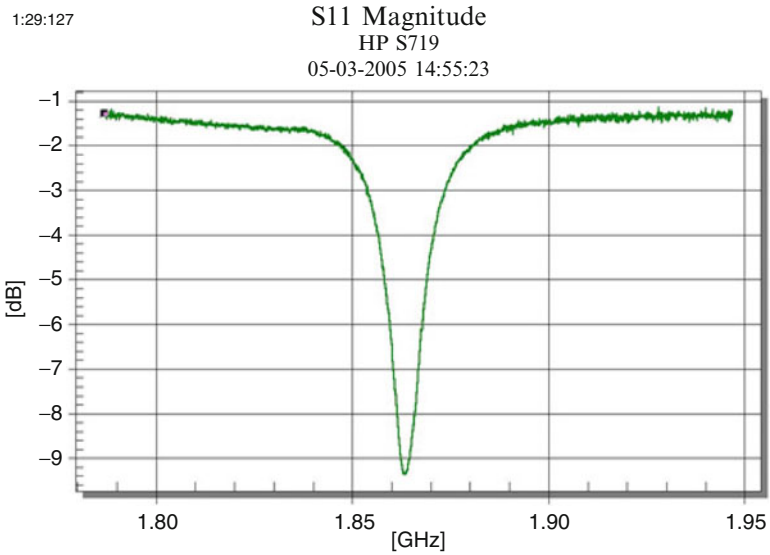

Fig. 9.9 Measured S11 magnitude for passive receiver

As a proof of concept for the BAW-based input matching circuit, the amplifier in Fig. 9.6 was replaced with a simple nonlinear envelope detector acting as an AM demodulator. Because no active gain is included in the receiver chain, the power consumption is extremely low for this "passive" receiver. To verify the matching topology, the receiver was implemented in 130-nm CMOS technology and a die photograph is shown in Fig. 9.8.

The BAW resonator is fabricated on a separate die and wirebonded to the CMOS circuitry. The measured sensitivity of the passive receiver was −38 dBm detecting on-off-keyed (OOK) data, with a power consumption of less than 1 μW. Figure 9.9 plots the measured S_{11} magnitude.

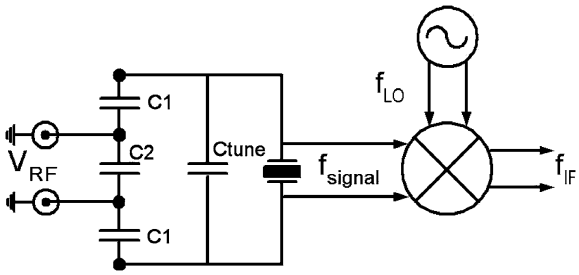

Fig. 9.10 Image-reject transformer architecture

9.2.3.2 Image-Reject Front-End Matching

This concept can be further extended to active receiver front-ends. When doing so, another interesting property of this inductorless, BAW-based impedance transformation emerges: a strong intrinsic image-rejection feature [3]. A fully differential realization of the transformer, along with a tuning capacitor and downconverter, is shown in Fig. 9.10.

Since the structure is now driven symmetrically, a virtual ground exists in the middle of capacitor C_2. Thus, only three transformer capacitors are needed. This is beneficial for the following reasons:

1. Capacitor C_2 does not have to be carefully matched on the positive and negative sides (although capacitor C_1 must be matched).
2. The capacitance value (and, thus, area) of C_2 can be reduced by a factor of two for a given transformation ratio compared to the single-ended version. Although C_1 is now duplicated, the value of C_2 is much larger than for upward impedance transformation.
3. Likewise, the tuning capacitance value (and area) can be reduced by a factor of two for the same tuning range due to the Miller multiplication inherent in driving the structure differentially.

One important property of this high-Q transformer is its ability to provide high levels of image rejection. At the series resonance frequency of the BAW resonator, the positive and negative nodes of the transformer are effectively shorted together through the low resonator impedance. At the parallel resonance frequency, however, the FBAR inhibits RF current from passing (recall that the ratio of the resonator impedance at the parallel and series resonances can be larger than 1,000). This dramatic impedance change over a very narrow range of frequencies can provide a large amount of image-rejection when used with an appropriate frequency plan. Referring to Fig. 9.10, we note that the intermediate frequency (IF) occurs at $f_{IF} = |f_{RF} - f_{LO}|$Hz. Thus, to achieve large image rejection ratios, one can place f_{LO} directly between the loaded series and parallel resonance frequencies of the resonator. The image will be located within a deep null provided by the series resonance, while the desired signal band exhibits gain (in the case of an active mixer). Figure 9.11 shows simulation results of the transformer transfer function.

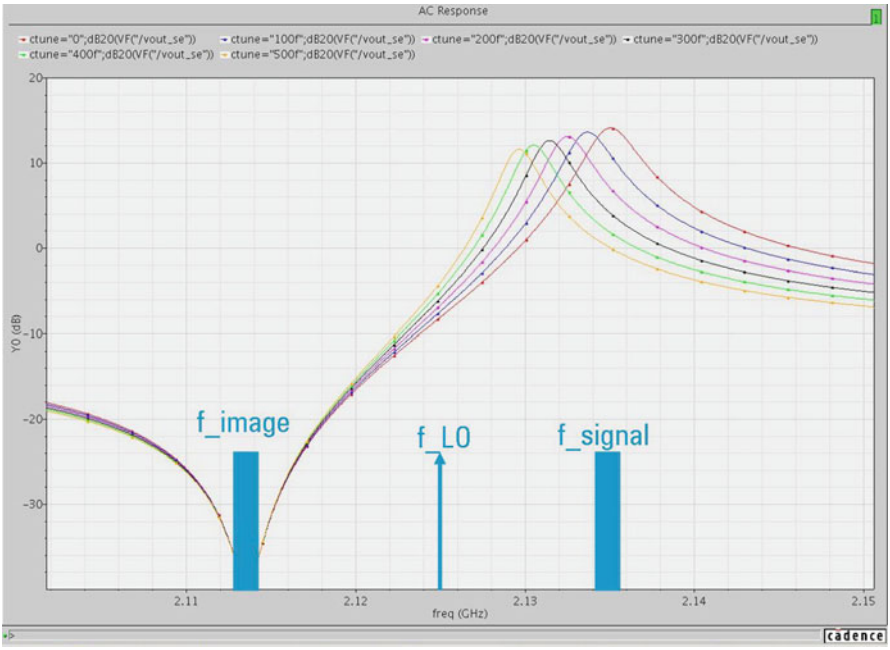

Fig. 9.11 Image-reject transformer transfer function. C_{tune} is swept to show transformer tuning capability

For this case, the impedance transformation ratio was approximately 20 (50–1,000 Ω). As expected, the transformer provides a very narrow passband due to the transformer quality factor (greater than 1,000). Since an upward impedance transformation of approximately 20 is performed, the expected voltage gain of the purely passive transformer is 13 dB. If the image-reject transformer drives the mixer directly from the antenna, this passive voltage gain helps suppress the large noise figures that low-power mixers typically exhibit. Capacitive tuning of the passband by varying C_{tune} is also shown in the simulation.

This technique requires only one mixer, no lossy filters, no quadrature generation, and can produce a relatively low intermediate frequency. One main drawback, however, is that the intermediate frequency is fixed (typically a few percent of the RF frequency, depending on the resonator coupling coefficient). If a fixed IF and a fixed RF bandwidth is acceptable, this is a useful technique for easily achieving 50 dB of image rejection.

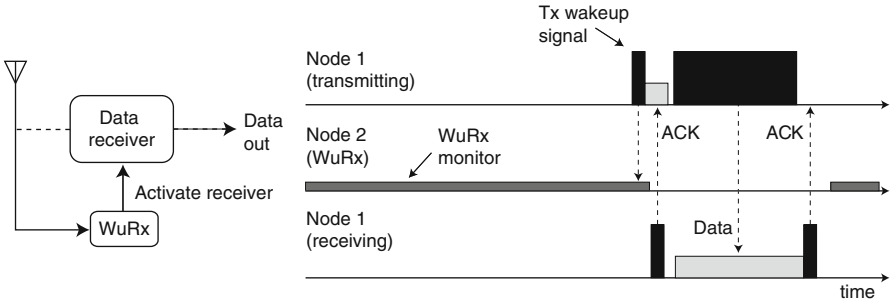

Fig. 9.12 Duty-cycle control with dedicated wake-up receiver

9.3 Two-Receiver Asynchronous Chipset

Combining the techniques introduced in Sect. 9.2 can lead to new and interesting transceiver architectures. We now discuss two new architectures that combine to form an on-demand wireless link with a 10-m range and an average power dissipation of 50 μW.

9.3.1 System-Level Considerations, Power Requirements

Wireless sensor network systems are characterized by very low duty cycles. Individual nodes spend most of the time in sleep mode, only activating the radio intermittently to communicate with neighboring nodes. A dedicated auxiliary receiver, or wake-up receiver, can be used to monitor the channel for communications and activate the main receiver when data is being transmitted. Because the wake-up receiver (WuRx) is listening all the time, the receiving node can react immediately to incoming requests. This eliminates the need for beaconing [4] and reduces network latency. The WuRx is not duty-cycled, however, so its power consumption must be extremely low. In most cases, the average power of the entire node is limited by the active power consumption of the WuRx.

A block diagram of the system is shown in Fig. 9.12. The WuRx sits alongside the main receiver and shares the antenna. Co-designing the WuRx and main receiver to use the same carrier frequency and modulation scheme ensures that both receivers operate with similar channel conditions. During sleep mode, the main receiver and the rest of the node's electronics are powered off, while the WuRx is active and listening for incoming requests. When the WuRx detects a predetermined sequence of bits, it activates the main receiver to handle data reception. The challenge for the WuRx design is to provide this listening functionality with a power consumption that is an order of magnitude or more below that of the main receiver. The power and

functionality of the two receivers may be optimized separately. The result is a chipset providing on-demand wireless connectivity with low latency and average power dissipation limited by the consumption of the WuRx.

9.3.2 Super-Regenerative Receiver Architecture

The past few years have seen a resurgence of the once-ubiquitous super-regenerative radio receiver. This architecture was first introduced by Armstrong in the 1920s [5]. Basic operation can be described as follows: the RF input signal is coupled into an oscillator tank. This oscillator is periodically enabled and disabled. When no input signal is present, the oscillator exhibits an exponential rise in amplitude as expected, where the initial condition is fundamental circuit noise. When an in-band RF signal is present, the oscillator starts up much faster since the exponential initial condition is now an RF signal. *The very high sensitivity of the oscillator's start-up time to the incoming RF signal is the source of the large RF gain achievable with this architecture.* Notice that this is a discrete-time sampling process, so signal and noise aliasing must be accounted for. Since analog audio reception was one of the primary applications of these receivers, the sample rate was chosen to be higher than the range of human hearing (supersonic) to avoid aliasing and audible tones, hence the name "super-regenerative." A subsequent demodulator circuit detects the pulse width of the oscillator envelope and averages over multiple "samples," effectively downconverting the signal to DC.

Reasons for its success during the first half of the twentieth century include:

1. It requires a minimal number of active devices, which was extremely important when vacuum electron tubes were the technology of choice. An entire receiver could be realized with two tubes: one RF oscillator tube and another detection tube.
2. Tremendous RF gain is realized without using an amplifier. This is probably the most intriguing aspect of a super-regenerative receiver.
3. The receiver can operate at extremely high frequencies (relative to the speed of the technology). This was a critical concern when using relatively slow vacuum tubes and is once again important if the RF devices are biased at low f_Ts to save power.

Although minimizing the number of active devices is clearly not a concern for modern RF design, the super-regenerative architecture is appealing due to its simplicity and potential for low-power dissipation. The receiver can operate above the f_T of the transistors, allowing subthreshold biasing that further reduces current consumption.

In traditional super-regenerative receivers, manual adjustments were required to fine-tune the oscillator frequency and loop gain [6]. Antenna reradiation, low spectral efficiency, distortion of analogue modulation, and the advent of

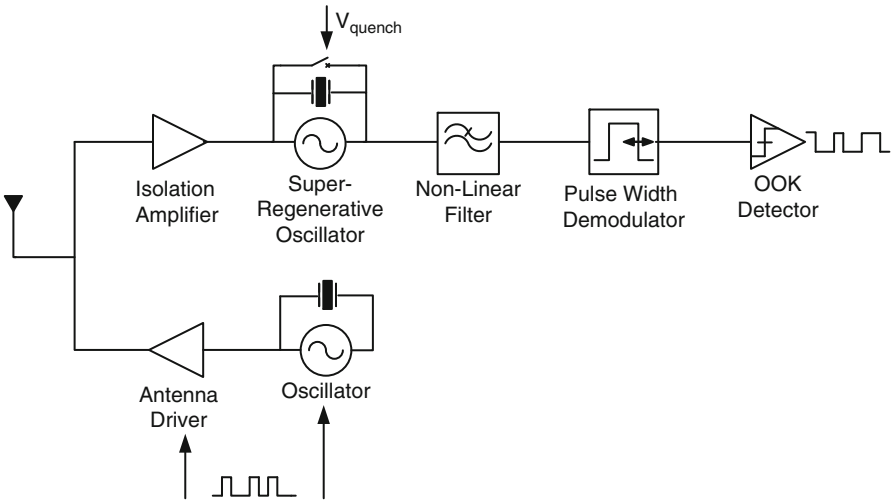

Fig. 9.13 Super-regenerative transceiver block diagram

inexpensive solid-state transistors allowing super-heterodyne architectures rendered the super-regenerative receiver obsolete by the late 1950s. However, we have shown that modern RF MEMS, CMOS technology, circuit design techniques, and digital communications solve many of the inherent problems with this architecture, providing very low power and high integration for sensor node applications [7].

Figure 9.13 shows the block diagram of the low-power transceiver.

The receiver consists of a passive matching network, an isolation amplifier, an amplifier with time-varying gain, and a bandpass positive feedback network forming an oscillator. The isolation amplifier between the antenna and the oscillator performs the following functions: it reduces RF leakage of the oscillation signal to the antenna, it provides an input match to the antenna via the passive matching network, and it injects the RF input signal current into the oscillator tank without adding significant loading to the oscillator. The time-varying nature of the loop gain is designed such that the oscillator transconductance periodically exceeds the critical g_m necessary to induce instability. Consequently, the oscillator periodically starts up and shuts off. The periodic shutdown of the oscillator is called "quenching." The start-up time of the oscillator is exponentially dependent upon the initial voltage of the oscillator tank. This dependency provides the large gain attainable by the super-regenerative receiver.

The RF signal is coupled into the oscillator tank from the antenna. It is then discretely sampled by the time-varying detector oscillator. Since this is a discrete-time sampled process, the quench rate must be at least twice the highest frequency component of the baseband signal. This quench tone must be filtered from the data signal to avoid corrupting the baseband signal. Thus, the oversampling ratio is determined by the complexity of this filter, and the data rate is ultimately limited by

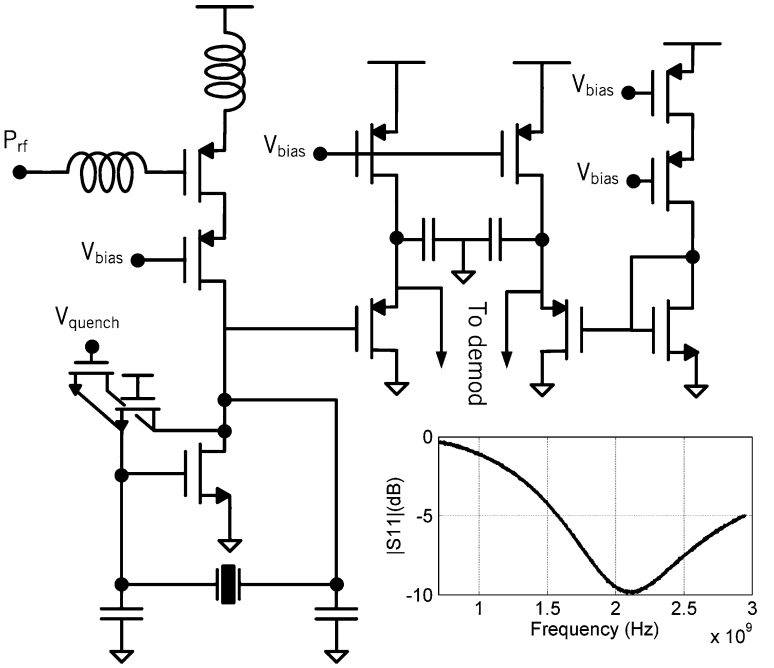

Fig. 9.14 Front-end schematic of super-regenerative receiver. The inset shows the measured receiver $|S_{11}|$

the filter bandwidth. However, the quench frequency cannot be arbitrarily increased to achieve a higher data rate. The frequency of the quench signal is limited by the time-varying growing exponential time constant of the oscillator, τ_{det}. For a high-Q resonant structure, such as a BAW resonator, τ_{det} is relatively long (greater than 100 ns). To allow a higher quench frequency, a short duty-cycled quench signal (e.g., 10%) can be utilized. In this implementation, such a short duty cycle is possible due to the very rapid quenching of the oscillator tank by the shunting transistor.

The front-end schematic is shown in Fig. 9.14.

The isolation amplifier comprises an inductively degenerated PMOS LNA with two on-chip inductors, yielding a fully integrated matching network. The most power-hungry components in the receiver—the isolation amplifier and detector oscillator—share their bias current, thereby effectively halving the current consumption of the receiver. The detector oscillator is cycled by the quench signal (V_{quench}), which creates a time-varying tank impedance, periodically dissipating the RF energy stored in the BAW resonator through a shunting transistor. The shunting transistor must be sized wide enough to quickly quench oscillation, but small enough to avoid loading the tank with unnecessary capacitance. A 50-Ω on-resistance is sufficient, yielding the addition of very low parasitic capacitance. The power consumed in

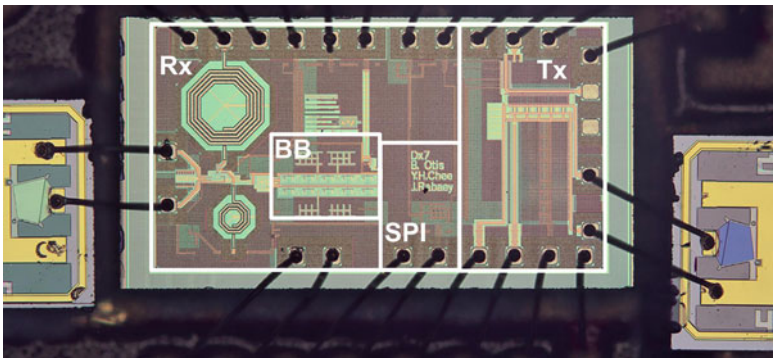

Fig. 9.15 Micrograph of the assembled integrated transceiver

driving the quench transistor must be absorbed into the receiver power budget. For a 300-fF capacitance (including pad and wiring capacitance), this power dissipation is approximately 30 nW. The shape of this impedance wave form allows for the tuning of the receiver gain and bandwidth properties. A square wave (10% duty cycle) impedance wave form allows simple and adaptable quench generation. Because the data symbols are oversampled, the exact quench frequency and phase is not crucial and may be readily supplied from a digital control block. The high-Q nature of the resonant BAW structure, providing a relatively long oscillator time constant and narrow intrinsic bandwidth, relaxes the need for precise control over the oscillator transconductance. Fine Tx/Rx frequency alignment can be achieved with relatively large, binary-weighted capacitor arrays due to the high-Q resonator.

A weak-inversion PMOS nonlinear filtering stage is DC-coupled to the oscillator. This stage consists of a PMOS source follower with a relatively low pole frequency (1 MHz). The 2-GHz carrier is attenuated, but the nonlinearity of the PMOS transistor creates a DC component proportional to the envelope of oscillation. By using a DC-coupled topology, the nonlinear bias point shift of the oscillator NMOS transistor adds to the PMOS nonlinear filter, thereby increasing the signal level to the pulse-width demodulator. A replica stage (with 20x current division relative to the super-regenerative core) was used to provide a pseudo-differential output of the nonlinear filter.

The on-off keyed (OOK) transmitter consists of a BAW-referenced oscillator to provide a stable RF carrier and an antenna driver to provide efficient power gain. Digital bits are directly modulated onto the carrier by on/off cycling the transmitter.

A fully integrated version of this architecture was fabricated in a 0.13-μm CMOS process (see Fig. 9.15 for a photograph of the assembled chip).

The transmit and receive resonators are clearly visible on the left and right sides of the chip, respectively. Standard commercial assembly techniques (chip placement with conductive epoxy and gold wirebonding) were used.

The total power consumption of the receiver when active is 380 μW. A majority of the power dissipation occurs in the RF front-end, where the required f_T of the

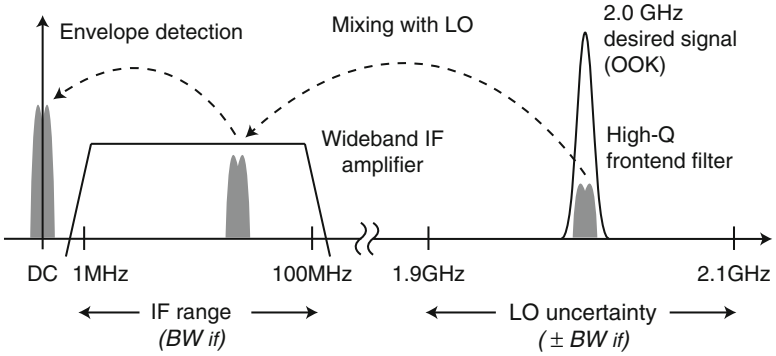

Fig. 9.16 Uncertain-IF receiver method of operation. ©2009 IEEE. Reprinted, with permission, from [8]

transistors is highest (around 2 GHz). The chip was tested with a quench frequency of 100 kHz and a 10% quench duty cycle. For a 10-kbps data rate, the measured sensitivity for a bit error rate (BER) of 10^{-3} is −97 dBm, displaying negligible degradation over a supply voltage range from 0.9–1.3V.

9.3.3 Uncertain-IF Chip

The super-regenerative receiver utilizes the MEMS resonator to stabilize the oscillator frequency. However, the high-Q nature of the BAW resonator is also useful as a filtering element. In this section, we describe a receiver architecture that takes advantage of MEMS-based frequency selectivity to relax the linearity and LO frequency accuracy requirements on the receiver circuitry [8]. The simplified receiver circuitry, enabled by the high-Q filter, results in substantial power savings. The concept is applied to the wake-up receiver, where the active power consumption is most critical.

Figure 9.16 shows the frequency plan and method of operation for the uncertain-IF receiver. First, the RF input signal passes through a front-end filter at the channel frequency. Next, it is mixed with an LO whose frequency is not precisely defined but instead is guaranteed to lie within a specified range ($\pm BW_{if}$). The resulting IF signal will then be located between DC and BW_{if}, where it can be amplified with greater power efficiency due to the low frequency. Finally, because the precise IF frequency is uncertain, an envelope detector performs the final downconversion to DC.

A block diagram of the implementation is depicted in Fig. 9.17.

The architecture relies on a sharp filter at the input to eliminate interferers, provided by the MEMS resonator. The filter allows the use of simple ring oscillator LO, which consumes about ten times less power than the power of the MEMS oscillator described in Sect. 9.2.1. The envelope detector is insensitive to variations

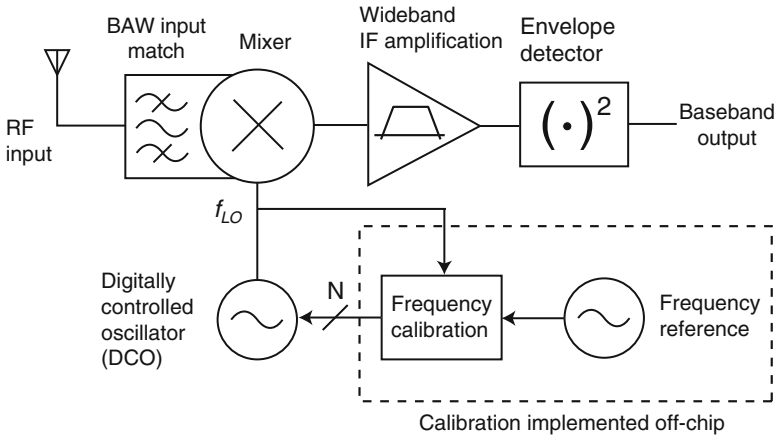

Fig. 9.17 Block diagram of uncertain-IF implementation. ©2009 IEEE. Reprinted, with permission, from [8]

in phase and frequency, so the phase-noise performance and frequency accuracy of the ring oscillator is much less important than in traditional architectures. The LO simply needs to be tuned to within BW_{if} of the RF channel frequency, which is easily accomplished using digital calibration techniques [9]. After calibration, the ring oscillator is allowed to free run, only requiring periodic re-centering to counteract aging or variations in temperature. Because the calibration operation is performed infrequently, its power consumption is amortized over time and can be neglected. The signal amplification is implemented at the lower IF frequency instead of RF, which provides more total gain through the receiver chain compared to an alternative implementation based on RF amplification [10].

A schematic of a prototype uncertain-IF implementation is shown in Fig. 9.18.

The input matching network employs the technique described Sect. 9.2.3 to implement a sharp RF channel filter. The signal then passes directly to the mixer, which is implemented with a single-ended dual-gate topology (M_1 and M_2). The single-ended design minimizes LO drive requirements because a differential LO signal is not needed; a single-ended CMOS ring uses half the power of a differential design. Any RF or LO feedthrough is suppressed by the load network and low-pass nature of the IF amplifier stages.

The IF amplifications is provided by a cascade of differential pair stages. The resistive loads require small voltage headroom and simplify interstage biasing. In order to mitigate accumulated offset throughout cascade of amplifiers, three of the stages are implemented using a split-source topology [11] to roll off the gain at DC. The envelope detector is a differential pair which exploits the nonlinearity of subthreshold MOSFETs M_3 and M_4 to convert the OOK signal to DC at the V_{bb} node.

Figure 9.19 shows a schematic of the digitally tunable ring oscillator.

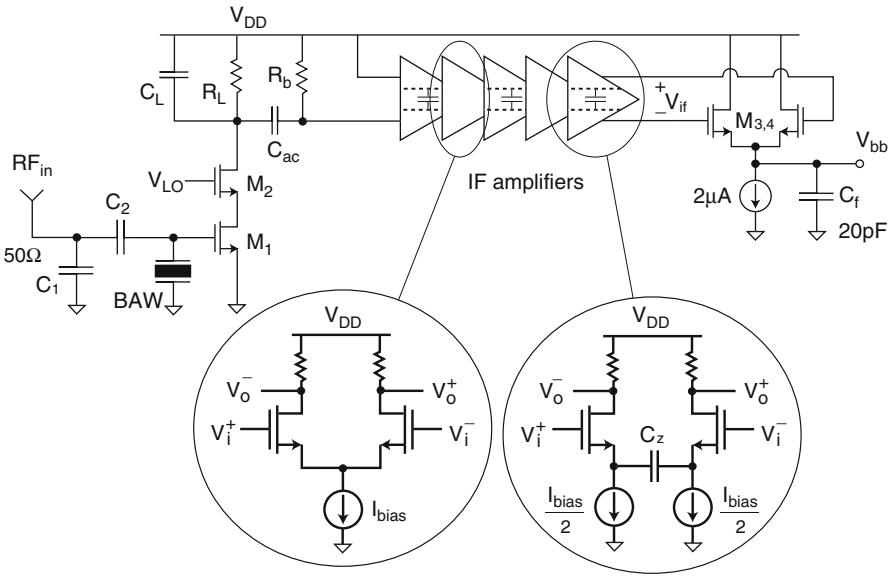

Fig. 9.18 Circuit-level schematic of wake-up receiver. ©2009 IEEE. Reprinted, with permission, from [8]

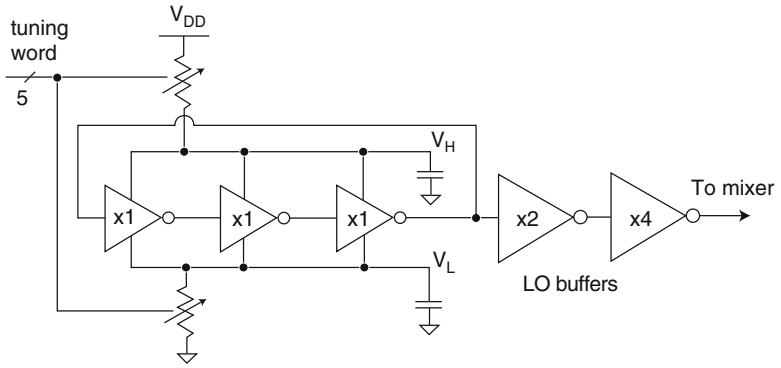

Fig. 9.19 Schematic of digitally tunable LO generator. ©2009 IEEE. Reprinted, with permission, from [8]

A simple three-stage CMOS ring is chosen to minimize the oscillator power consumption. Recall that LO phase noise and frequency stability are not critical for the uncertain-IF architecture. Frequency calibration is enabled through the digitally tunable resistors that modify the virtual supply rails V_H/V_L of the ring. A scaled inverter chain restores rail-to-rail output swing and drives the LO port of the mixer (gate of M_2 in Fig. 9.18).

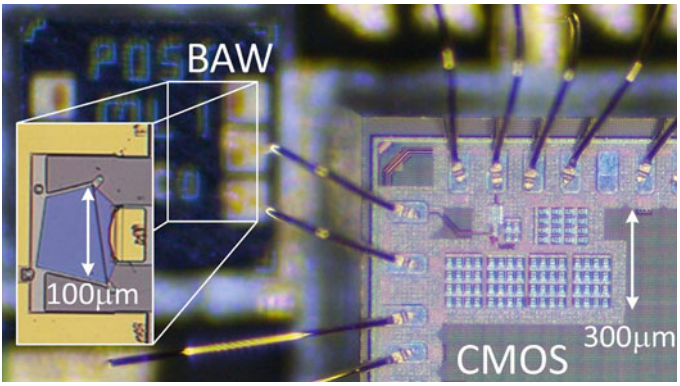

Fig. 9.20 Die photo of uncertain-IF receiver prototype

The prototype receiver was implemented in 90-nm CMOS technology. The die photo is shown in Fig. 9.20.

The packaged BAW can be seen on the left side of the CMOS die; the inset shows the location of the active resonator structure. Because the high-Q resonator is simultaneously used for input matching and channel filtering, no additional on-chip inductors or resonators are required. The result is a compact implementation and small CMOS area.

In order to reduce power dissipation and take advantage of the low threshold voltages in this deep submicron technology, the entire receiver is designed to operate from a 0.5-V supply. The measured normalized gain is plotted in Fig. 9.21 for four different samples. The high-Q resonance of the BAW filter is evident in the frequency response.

For each sample, a calibration is first performed to tune the ring oscillator LO frequency within BW_{if} from the channel frequency defined by the BAW resonator. Therefore, each sample will have slightly different LO frequency due to natural process variation. The calibrated LO frequencies are also marked in Fig. 9.21. In spite of the variability in the LO frequency, each sample is able to detect a signal at the RF channel frequency due to the wide IF bandwidth. In effect, the burden of selectivity has been shifted from the LO to the front-end filter provided by the BAW resonator. The measured receiver sensitivity is -72 dBm while receiving OOK data at 100 kbps with a BER of 10^{-3}. The power consumption of the WuRx prototype is just 52 µW from the 0.5-V supply. Of the total, about 20 µW is dedicated to the CMOS ring oscillator and LO buffers running at 2 GHz. The BAW channel filter is the key component that allows the use of a free-running ring oscillator as the LO, despite the poor phase noise and frequency accuracy of this type of oscillator.

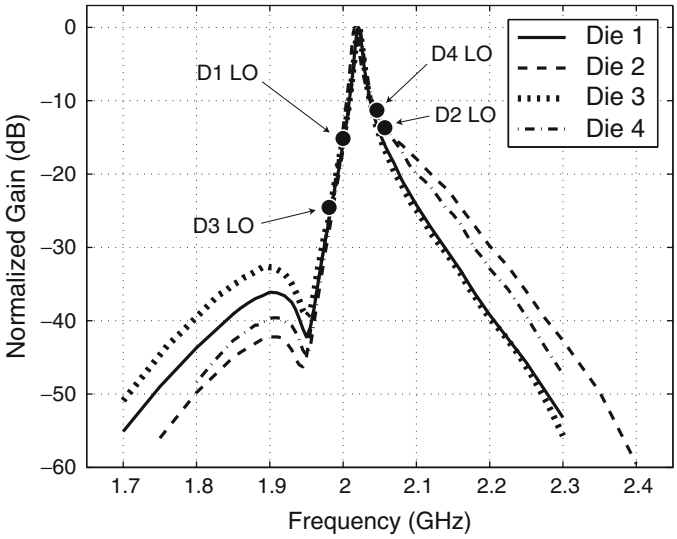

Fig. 9.21 Normalized gain response for four different samples

9.4 Conclusions/Toward the Future

In the past few decades, a huge body of research has enabled the realization of fully integrated transceivers using very poor quality passive components (on-chip inductors, capacitors, and a front-end filter). To accomplish this, the burden of frequency definition has been placed on power-hungry active circuitry. For example, to create a low phase-noise LO using an on-chip LC tank, the designer would likely maximize the VCO signal swings and maximize the PLL loop bandwidth, at the expense of power and reference spurs, respectively. In contrast, the overarching theme of this chapter is the following: push the burden of frequency definition into the passive domain as much as possible. We have shown three examples of this strategy:

1. Oscillator tuning using BAW resonators
2. High-Q band-select amplifier tuning
3. Passive image-reject impedance transformation

As MEMS devices improve and become more widely available, circuit designers will continue to leverage their unique properties. Future work will likely include the design of low-jitter frequency references, bandpass oversampled data converters using high-Q MEMS resonators, and temperature-compensated frequency references aimed at quartz-replacement.

Acknowledgment The authors would like to acknowledge the support of Avago Technologies and the BWRC member companies.

References

1. I. Larson, J.D.P. Bradley, S. Wartenberg, R. Ruby, 2000 IEEE Ultrason. Symp. **1**, 863 (2000)
2. B. Otis, J. Rabaey, IEEE J. Solid State Circ. **38**(7), 1271 (2003)
3. B.P. Otis, Frequency-selective transformer and mixer incorporating same. U.S. Patent Publication 2007/0205849, 2006
4. E.Y. Lin, J. Rabaey, A. Wolisz, IEEE Int. Conf. Commun. **7**, 3769 (2004)
5. E. Armstrong, Proc. IRE **10**, 244 (1922)
6. J. Whitehead, *Super-Regenerative Receivers* (Cambridge University Press, Cambridge, 1950)
7. B. Otis, Y. Chee, J. Rabaey, IEEE ISSCC Digest of Technical Papers, p. 396 (2005)
8. N. Pletcher, S. Gambini, J. Rabaey, IEEE J. Solid State Circ. **44**(1), 269 (2009). DOI 10.1109/JSSC.2008.2007438
9. T.H. Lin, Y.J. Lai, IEEE J. Solid State Circ. **42**(5), 1003 (2007)
10. N. Pletcher, S. Gambini, J. Rabaey, in *Proceedings of IEEE Custom Integrated Circuits Conference*, 2007, pp. 539–542
11. D. Daly, A. Chandrakasan, IEEE J. Solid State Circ. **42**(5), 1003 (2007)

Chapter 10
A 2.4-GHz Narrowband MEMS-Based Radio

David Ruffieux, Jérémie Chabloz, Matteo Contaldo, and Christian C. Enz

Abstract This chapter presents an innovative wireless transceiver architecture that rely on MEMS components to achieve further miniaturization and significant power dissipation reduction compared to low-power radios targeting LDR to MDR applications. It is shown in particular how the limitations of MEMS devices can be waived at the architectural level and how their combination can lead to innovative concepts preserving or even surpassing the performances of current mainstream optimized solutions. Besides the architectural aspects, the chapter also focuses on the design of some ultra-low-power and MEMS-specific circuits and reports measurement results of the complete system. The synthesizer, which is based on a low-phase-noise fixed-frequency BAW DCO and a variable IF LO obtained by fractional division from the RF carrier, achieves a phase noise of −113 dBc/Hz at 3 MHz. To correct for its ageing and thermal drift, the BAW DCO can intermittently be phase locked to a 3-μA, ±5-ppm, 32-kHz reference, which is obtained after temperature-dependent fractional division of the signal of a 1-MHz silicon resonator so as to compensate the non-idealities of the latter (frequency tolerance, large thermal drift). An all-digital PLL implementation guaranties a nearly immediate synthesizer settling when returning from an idle period, owing to the memorization of the previous lock conditions eliminating a multi-MHz XTAL and its associated start-up time. A sensitivity of 87 dBm was obtained in receive mode at 100 kb/s for a global consumption of 6 mA. The transmitter demonstrates a high-data-rate quasi-direct 1-point modulation capability with the generation of a 4-dBm, 1-Mbps, GFSK signal with an overall current of 20 mA. Both the receiver and transmitter further take advantage of BAW filters to implement interferers, image, and spurious rejection.

D. Ruffieux (✉) • J. Chabloz • M. Contaldo • C. C. Enz
CSEM, Centre Suisse d'Electronique et de Microtechnique, Neuchâtel, Switzerland
e-mail: david.ruffieux@csem.ch

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_10,

10.1 Introduction

Miniaturization and reduction of the power dissipation are the main issues that currently prevent the seamless integration of wireless and networking capability into any tiny device used in short distance wireless applications. The latter include low-data-rate (LDR) (typically 10 kb/s) applications such as wireless sensor networks (WSN), wireless body area networks (WBAN) but also medium-datarate (MDR) (typically 1 Mb/s) wireless applications such as audio streaming, wireless medical devices, wireless miniature drug delivery systems, or wireless implants. Today, both LDR and MDR use different radios that are individually optimized for each application. Nevertheless, these applications are now converging, and it is the main objective of this work to design a single miniaturized radio covering the entire spectrum of short distance wireless applications from LDR to MDR.

In short-range wireless communication, the link budget is often comfortable resulting in an excess signal-to-noise ratio (SNR). When available, such excess SNR can advantageously be traded for higher bandwidth to boost the data rate and decrease the radio duty cycling. This is a promising way to increase the autonomy of battery-operated nodes provided the quartz start-up and PLL settling times can be reduced or even eliminated.

In many short-range wireless applications requiring an ultra miniaturized radio, the SAW filters and the crystal quartz remain the bulkiest parts. For hearing aids, for example, this prevents the introduction of a radio link into the ear canal. The system miniaturization can be addressed first by a higher integration with a system on-chip (SoC) embedding the radio chip together with other required functions (sensor interface, μC, memory, power management,…) as demonstrated in [1] and second with a system-in-package (SiP) approach to include all passive components (RF SAW filters, impedance matching networks, quartz crystal,…). The external SAW filters can then advantageously be replaced by bulk acoustic wave (BAW) filters implemented either on top of the IC with an above-IC approach [2] or in a SiP [3]. Similarly, quartz crystal unit(s) can be replaced by miniature silicon resonator(s).

10.2 Duty Cycling for Long Node Autonomy and Current Limitations

Wireless sensor nodes are faced with tremendous power concerns that are further exacerbated as node miniaturization is desired. A 1-cm^3 lithium battery (e.g., CR2032) can deliver some 200 mAh. Assuming a 10- mA radio, it translates to no more than 20 h of continuous operation. The only way to ensure longer autonomy is to duty cycle the radio activity calling for efficient synchronization techniques to avoid wasting power needlessly. Nonetheless, many applications, such as remote control, require low latency, meaning that the link must set up very fast should an event be communicated after a long idle period. For example, a latency of 100 ms is nearly unnoticeable by human beings.

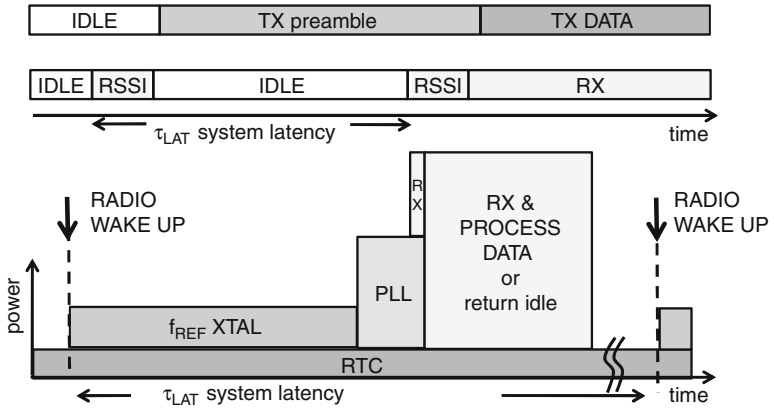

Fig. 10.1 Illustration of CSMA with preamble sampling (*top*) and associated wake-up sequence with impact on power dissipation (*bottom*)

The top part of Fig. 10.1 illustrates a state-of-the-art low-power radio protocol adapted to sensor nodes applications and known as carrier sense multiple access (CSMA) with preamble sampling [4] that is particularly suited for rare signaling traffic and for detecting incoming nodes. In this scheme, the receivers of the destination or relay nodes are mostly idle but wake up at regular intervals according to the desired latency of the system (reaction time) to listen if an activity is detected on a predetermined channel with a receive signal strength indicator (RSSI). A node having information to transmit only needs to emit a preamble that is slightly longer than the system latency before actually sending data to make sure the receiver can detect it despite the receiver/transmitter lack of synchronization.

Once aware of each other, such nodes could exchange regular data packets in a synchronized way using other channels and spatial time division multiple access TDMA [4]. The maximum idle time duration between packets in TDMA is limited by the accuracy of the nodes real-time clocks (RTC). The receiving node has to start listening earlier and longer to make sure it can account for RTC mismatches. Assuming an overall timing accuracy of 100 ppm for any node, hence 200 ppm maximum deviation, among any two nodes synchronization can be maintained over some 8 min before the added listening time reaches 100 ms again and from where it would be appropriate to switch back to unsynchronized mode. If the nodes exchange data every minute, the extra listening time can be reduced to 12.5 ms while suppressing the TX preamble in this example.

The bottom part of Fig. 10.1 illustrates the wake-up sequence and associated power dissipation when sensing the communication channel in CSMA mode. In such systems, when sleep mode is activated, only the RTC is maintained, calling for very low-consumption implementation (1 μA) that can only be achieved with a wristwatch 32- kHz crystal if significant RTC accuracy is desired. On the other hand,

a conventional transceiver system will require a more precise, higher frequency radio XTAL to serve as a low noise reference for the radio operation. Duty cycling of the latter is mandatory to lower the average dissipation, but it translates into a rather long start-up time close to 1ms during which a lot of current is required to accumulate mechanical energy in the resonator. When the quartz oscillator frequency and amplitude reach steady state, the phase-locked loops (PLL) can be started to provide the desired synthesized RF carrier after an additional settling time of about 200 µs (assuming a 10- kHz bandwidth). Only then the receiver can be turned on to deliver after averaging over a few symbols an RSSI level and decide whether to return to sleep mode or to keep listening. This translates to a period of about 1.2 ms during which the system is burning power needlessly since no information can be retrieved during that time. Furthermore, it limits the latency to above 100 ms if duty cycling in the range of 1–2% is desired to significantly reduce the average dissipation.

Eliminating the crystal start-up and PLL settling times would allow achieving significant power savings not only by sparing the associated energy but mostly by enabling short burst of data to be transmitted at higher data rate with reduced TDMA time slots. A 10-kbps average throughput could, for example, be maintained for some periods with 1–100% duty cycling and peak data rate between 1 Mbps and 10 kbps depending on the link quality. Since the power consumption of a transceiver system does not significantly depend on the data-rate for LDR to MDR applications, trading excess link margin for bandwidth and lower cycling could result in true power scaling. A frequency agile synthesizer would further allow a rapid sampling of the different channels for increased robustness (e.g., redundant CSMA channels), ease the implementation of efficient frequency hopping schemes, and get dynamically an estimate of the load of the different channels for improved coexistence.

10.3 Where to Use MEMS Components and How to Waive Their Limitations

Besides the inherent size reduction that could be obtained with the replacement of external parts of a transceiver with MEMS counterparts, one should seek for ways to further reduce the power dissipation. At individual block level, using RF MEMS filters in the receiver and RF MEMS resonator to implement the local oscillator offers such opportunities. BAW resonators can be used in the RF front-end taking advantage of their high-Q for interferers rejection so as to relax the receiver global linearity requirements, thus lowering the power consumption. With a superheterodyne architecture using a double-sideband (DSB) mixer and relying on the filter selectivity to reject the image, the generation of quadrature LO signals can be avoided again lowering the battery load. BAW resonators can obviously

also be used in the oscillator to achieve extremely low phase noise at low-power consumption [5–9]. One of their main drawbacks is their reduced tuning range inherent to their small coupling coefficient. One approach used recently to extend the tuning range of BAW oscillators is to use a negative active capacitance to tune out the BAW parallel capacitance [10]. Unfortunately, this comes at the cost of a significant reduction of the Q factor. An interesting alternative is to handle the reduced tunability of the BAW oscillator at the architectural level and shift the channel selection issue to intermediate frequency with the introduction of a wide-IF concept as described in the next section.

Silicon MEMS resonators, which have recently entered the market trying to rival with quartz crystals in consumer applications, are attractive candidates for the replacement of the latter. More compact than their quartz counterpart and cost effective, thanks to mass production and wafer scale whole silicon encapsulation, they however suffer from a high temperature coefficient of frequency (TCF) in the range of -30 ppm/K dominated by the variation of the elastic properties of silicon. Structural techniques have been proposed for first-order TCF cancelation by balancing the silicon stiffness variation with e.g., the introduction of temperature-dependent axial stress into the resonator beam [11], applying an electric field-induced stiffness control with temperature-dependent variable gap [12] or using composite beams combining materials with positive (silicon dioxide) and negative thermal stiffness coefficients [13, 14]. While interesting, such concepts tend to significantly complicate the resonator fabrication process impacting cost, robustness, and reliability or require precise thickness ratio control raising uniformity and reproducibility issues. Furthermore, they only address part of the overall problem by considering solely the resonator itself and trying to mimic quartz parts which have such properties for granted. In particular, these approaches miss the generalization that is demanded by large-scale manufacturing and that calls for a few standardized designs and other means to implement customization via programmability. Avoiding moreover mechanical frequency trimming of individual resonator as it is done with quartz parts to compensate fabrication tolerance is as important when considering wafer-level encapsulation.

Electronic compensation based on fractional PLL synthesizer techniques elegantly solved all of the above issues simultaneously [15], in contrary to applying a temperature-dependent bias to the resonator as proposed in [16], which is barely sufficient to address temperature and tolerance compensation. One of the main drawbacks of the PLL approach is the increased power consumption that makes such a solution unpractical for low-power RTC that should achieve μW level dissipation. The concept that will be illustrated below bridges the gap with the demonstration of a generic ultra-low-power temperature compensation concept applied to a low-frequency uncompensated piezoelectric AlN-driven silicon resonator [14] that can be used to implement both an RTC [17] and a reference for a MEMS-based frequency synthesizer architecture [18] eliminating the need for any other bulky reference.

10.4 MEMS-Based Radio Architecture

A new heterodyne radio architecture is proposed that basically keeps a fixed-frequency high-Q BAW RF oscillator and shifts the tunability problem to the intermediate frequency (IF), preserving the low phase noise of the BAW oscillator and relaxing the phase-noise requirement on the IF PLL, thanks to frequency scaling. This requires a wide tuning range at IF achieved, thanks to a new relaxation oscillator. This approach also allows merging the reference frequency and timekeeping functions into a single low-power and low-frequency oscillator, using a low frequency resonator, hence eliminating a radio crystal and shortening drastically the radio turn-on time.

Figure 10.2 depicts the architecture of the proposed MEMS-based superheterodyne transceiver designed in the context of a miniature radio solution for the 2.4-GHz ISM band [18]. The main local oscillator is a fixed 2.32-GHz or 2.56-GHz RF LO based on a high-Q BAW resonator yielding 30-dB improvement in the phase-noise power product compared to LC implementation [9]. Such BAW oscillators hence do not require being placed within high-bandwidth multi-MHz quartz referenced PLL to achieve acceptable close-to-carrier noise levels. While ageing and high thermal coefficients of RF BAW resonators remain a serious issue for stable frequency generation, the large time constants associated to such phenomenon allow using a PLL with a much lower reference frequency and loop bandwidth for their compensation so that the increased noise of the reference can also be suppressed. Merging the reference frequency and timekeeping functions into a single low-power low-frequency oscillator spares another bulky radio XTAL further eliminating the associated start-up time, while maintaining low-power dissipation when the radio is idle and only the RTC is running. To ensure a rapid settling time upon radio wake-up despite the low loop filter bandwidth, a simple all-digital PLL (ADPLL) [19] has been implemented so that exact previous lock state

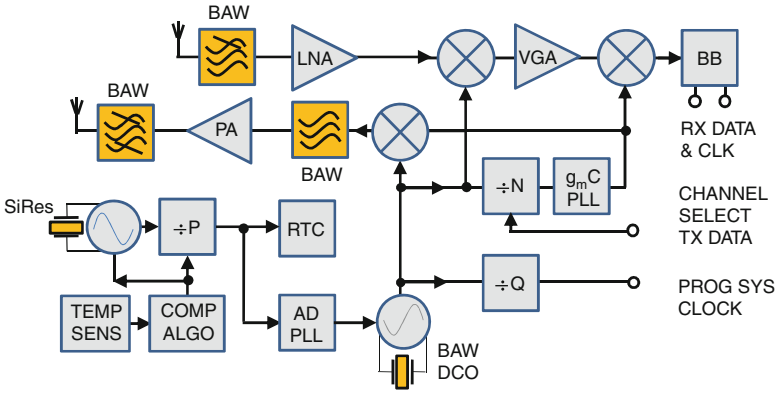

Fig. 10.2 Architecture of the narrowband, multichannel MEMS-based radio

conditions can be memorized and the PLL started almost locked. The phase error is quantized on a time-to-digital converter (TDC) and filtered numerically before it switches the varactor bank of the digitally controlled BAW oscillator (DCO). The poor tuning capability of BAW resonators is thus turned into an asset to yield a DCO with a very fine frequency resolution and high noise immunity. A 32-kHz time and frequency reference is obtained from a higher frequency silicon resonator after temperature-dependent fractional division. Further details on the concept and the implementation of the compensated silicon resonator are given in the following section.

In addition to the fixed RF carrier, the proposed RX and TX superheterodyne architecture requires quadrature signals from 80 to 160 MHz at intermediate frequency (IF) to cover the whole ISM band. These signals are derived from the fixed RF carrier by an open-loop fractional division chain by 13–35 with inherent digital modulation capabilities. The required BAW LO tolerance can then be derived and yields 0.65% with a low side injection, while it is relaxed to 0.9% in a high side injection implementation. Such tolerances are similar to what is needed for BAW filter implementation and can be attained in volume production with wafer-level trimming equipment [20]. A homodyne relaxation VCO PLL is then used to remove the far-from-carrier quantization noise due to the $\Delta - \Sigma$ modulator noise shaping and produce quadrature outputs [21]. The bandwidth should be chosen wide enough to cancel out the large close-to-carrier noise of the relaxation oscillator, reduce settling time, and allow in-band modulation at up to 2 Mbit/s yielding a calibration-free quasi-direct digital modulation transmitter architecture suitable for standards such as Bluetooth low energy [22] or Zigbee [23].

The receiver front-end consisting of a differential LNA and a double-sideband mixer is preceded by a lattice BAW filter to relax the front-end linearity requirements and yields a 50-dB image rejection without requiring RF quadrature signals [24]. A variable gain amplifier is provided at IF before I&Q direct down-conversion. Channel selection is performed by a few passive poles and a third-order Butterworth Gm-C filter. The signals are then fed to two limiter chains to provide RSSI information and rail-to-rail I&Q data.

The transmitter section consists of a DSB up-converter mixer combining the RF and filtered modulated IF signals that is loaded by a single-stage BAW filter to suppress the unwanted sideband before the wanted signal is amplified by a preamplifier plus power amplifier chain and filtered again. Besides avoiding power-consuming quadrature RF LO signal generation, this approach further allows distributing the filtering stages, hence reducing the insertion loss after the power amplifier. Inductive resonant loads where chosen both at the mixer output and preamplifier input to provide well-defined real impedances for the intermediate BAW filter that was optimized for 600-Ω loads.

A 16-MHz system clock is derived by fractional division from the RF carrier to assist data modulation/demodulation. These modem functions are currently implemented on an FPGA for greater flexibility, thanks to a parallel bus interface.

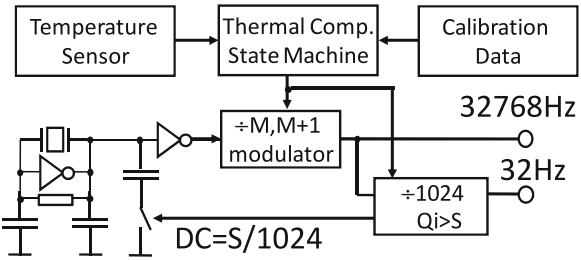

Fig. 10.3 Temperature-compensated clock block diagram suitable for silicon resonator

10.4.1 Low-Power Electronic Compensation of Silicon Resonator Imperfections

Figure 10.3 shows the block diagram of a temperature compensated clock that is based on a 1-MHz AlN-driven silicon MEMS resonator and that produces an accurate 32,768-Hz clock at a power dissipation level suitable for RTC applications. This architecture eliminates the power-hungry PLL without compromising any of its advantages such as programmability, temperature, and fabrication tolerance compensation.

It relies on frequency interpolation, which is obtained by switching dynamically the load capacitance of the resonator between two values with a variable duty cycle (dc) for fine-tuning and the introduction of programmable division (N) for extended temperature range compensation and initial frequency adjustment. The relative stability of the divided clock, df_{div}^{rel} as a function of the parameters introduced above (N, dc), the relative frequency difference between high and low-frequency modes Δf_{HL}^{rel}, and a variation of the temperature ΔT can be written as

$$df_{div}^{rel}(\Delta T, N, dc) = \frac{\frac{f_o}{N} \cdot (1 + \alpha \cdot \Delta T + \beta \cdot \Delta T^2 + dc \cdot \Delta f_{HL}^{rel})}{\frac{f_o}{N_o} \cdot \left(1 + dc_o \cdot \Delta f_{HL}^{rel}\right)} - 1, \tag{10.1}$$

where α, β are the first- and second-order TCF of the resonator, f_o, its nominal frequency in low-frequency mode, and N_o, dc_o chosen to compensate its fabrication tolerance at calibration and obtain the desired divided frequency. Setting (10.1) to zero and solving for $0 \leq dc(\Delta T, N) \leq 1$ by adjusting N yield the condition for temperature compensation. With Δf_{HL}^{rel} in the range of 50 ppm, N needs to be greater than 20'000. With f_o close to 1 MHz, the compensated frequency must lie below 50 Hz.

To overcome this limitation and yield programmability, division by N is obtained by cascading a dual modulus divider $(M, M+1)$ controlled by a first-order modulator together with a fixed divider by a power of two (1,024 in the example). When the modulator and fixed divider use similar number of bits, programmable integer division variable by unity can be obtained $(1,024 \cdot M \leq N \leq 1,024 \cdot (M+1))$. The resolution of the bi-frequency mode depends on the fixed divider size whose output

signal duty cycle can be varied by detecting any intermediate state of the counter ($Qi > S$). The overall division ratio is chosen big enough to guaranty the continuity of the temperature compensation but at a scaled power of two of the desired output clock (e.g., 32 Hz and 32,768 Hz in RTC applications). With a bi-frequency range close to 50 ppm, an interpolated 32,768-Hz clock that is accurate 32 times per second and that can be adjusted with a resolution of 50 ppb can hence be obtained.

To maintain the power dissipation low, only the oscillator and divider chain are powered permanently. Periodically, the updated division ratio and bi-frequency duty cycle are calculated with the help of (10.1) on a dedicated hardware state machine or a microcontroller using the output of a temperature sensor and stored calibration data obtained at three temperatures from both the thermal sensor and the resonator. These blocks are heavily duty-cycled to yield negligible power overheads.

10.5 Frequency Synthesizer

10.5.1 BAW DCO Within ADPLL

The differential oscillator structure that has been chosen to implement the BAW RF LO was proposed in [25] and is sketched in Fig. 10.4. The DC decoupling of the two sources of the cross-coupled pair is necessary to avoid latching, and let the bottom MOS transistors set the oscillator common-mode voltage. This negative feedback is canceled above a given frequency and changes to positive to provide the required negative conductance at the desired frequency. The upper value of C_s is however limited so that parasitic relaxation oscillation can be avoided [25].

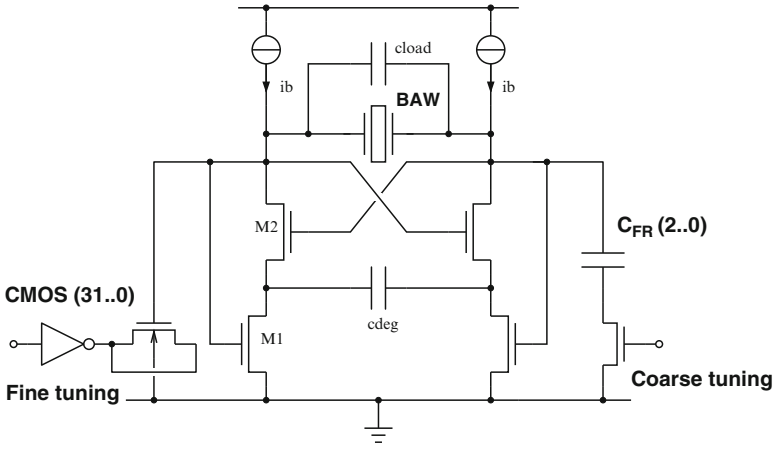

Fig. 10.4 BAW DCO schematic (coarse and fine-tuning both implemented differentially). ©2009 IEEE. Reprinted, with permission, from [27]

The resonator can be modeled by its dielectric capacitor C_0 placed in parallel with a motional branch made of a series RLC network that represents the resonance mode of interest and for which a suffix "m" is added in the following. The equivalent loading capacitor C_L, placed in parallel with the resonator, represents the capacitive load of the LO and the blocks to which it is connected. It could include an optional varactor if tuning is desired. The equivalent parallel conductance G_P of the tank that the circuit will have to compensate for is given by [25]

$$G_P = \frac{\omega_0 \cdot C_0}{Q \cdot \frac{C_m}{C_0}} \cdot \left[1 + \frac{C_L}{C_0}\right]^2, \tag{10.2}$$

where Q is the loaded Q factor to be derived shortly and ω_0 the oscillator frequency that is given by

$$\omega_0 = \frac{1}{\sqrt{L_m \cdot C_m}} \cdot \left[1 + \frac{C_m}{2 \cdot (C_0 + C_L)}\right]. \tag{10.3}$$

The oscillator total start-up current can readily be obtained after calculating $g_m(I_b)$, which is equated to $2G_P$, and doubled. It can be minimized if the MOS are biased in weak inversion and the load capacitance is reduced. When the circuit is supplied with a bigger current, the differential steady-state oscillation amplitude V_{OSC} depends on the nonlinearity of the MOS and can be found by solving $I_{f0}/V_{OSC} = G_P$, with I_{f0} the amplitude of the differential current at the fundamental oscillation frequency f_0 [26]. When expressed as a function of the mean total current I_M, the amplitude is bounded by $V_{OSC} < I_M/G_P$. The equality can however be used to provide a fairly accurate $V - I$ behavior of the LO amplitudes above a few hundreds of mV. The phase noise, L, of the oscillator when considering only thermal white noise can be computed as a function of the resonator and circuit parameters. With k the Boltzmann constant, T the temperature, and γ the noise factor of the oscillator circuit, it yields

$$L(\Delta\omega) = \frac{(1+\gamma) \cdot k \cdot T}{\omega_0 \cdot C_0 \cdot Q \cdot \frac{C_0}{C_m} \cdot \left(1 + \frac{C_L}{C_0}\right)^2} \cdot \frac{1}{V_{OSC}^2} \cdot \frac{\omega_0^2}{\Delta\omega^2}. \tag{10.4}$$

When considering a low-power implementation, a relevant FOM is obtained by computing the power phase-noise product, which should be minimized (e.g., obtaining a given phase noise while minimizing the power dissipation). As for LC tanks, it turns out to be only dependent on $1/Q^2$, V_{DD}/V_{OSC}, and the oscillator noise factor and shows no dependency on any reactive component of the tank.

The loaded Q is given as a function of the resonator and load Q factors, $Q_R = 1/(\omega_0 \cdot R_m \cdot C_m)$ and $Q_L = (g_0 + g_L)/(\omega_0 \cdot (C_0 + C_L))$, respectively, by

$$\frac{1}{Q} = \frac{1}{Q_R} + \frac{C_m}{C_0 + C_L} \cdot \frac{1}{Q_L}. \tag{10.5}$$

Interestingly, a moderate load Q factor, which also accounts for the resonator dielectric loss (g_0) is boosted by the tapping effect of the resonator (multiplied by $(C_0 + C_L)/C_m$). As an example, with $Q_L = 25$, $C_m/(C_0 + C_L) = 2.5\%$, and $Q_R = 1{,}000$, the loaded Q is still 500, yielding a 34-dB improvement in the power phase-noise product compared to an LC tank with a loaded Q of 10 (assuming same amplitude, supply voltage, and noise factor). The drawback, however, is the extremely reduced tuning range of the BAW LO which can be evaluated by computing the derivative of the oscillation frequency (10.3) with respect to a variation of C_L. It yields

$$\frac{\Delta\omega}{\omega_0} = \frac{C_m}{C_0} \cdot \frac{C_L}{C_0} \left[1 + \frac{C_L}{C_0}\right]^{-2} \cdot \frac{\Delta C_L}{2 \cdot C_L}. \tag{10.6}$$

It is scaled at best by $1/4 \cdot C_m/C_0$ if $C_L = C_0$ compared to an LC tank. Hence, it is close to a factor hundred with the coupling factor given previously. Advantage can be taken from this poor tuning behavior to implement a high-frequency resolution DCO that will prove very useful to ensure nearly immediate locking after returning from an idle period or compensate a difference of pulling between RX/TX modes. Figure 10.4 shows the two kinds of discrete varactors that have been used to implement a bank of 32 4×0.18-μm^2 inversion/depletion nMOS capacitances for fine tuning and a bank of 3 coarse-tuning switched metal–metal fingered capacitances. Sub-ppm frequency steps are obtained with frequency interpolation by switching four fine-tuning MOS varactors with a 6-bits MASH 1–1 modulator as proposed in [19].

The ADPLL implementation is based on a programmable integer divider, a phase frequency detector that is followed by a 100-ps resolution time-to-digital converter (TDC) and a simple proportional/integer digital loop filter. Self-calibrating compensation of the deterministic jitter introduced by the $\Delta-\Sigma$ modulator and bi-frequency operation when generating the reference signal for the ADPLL was incorporated within the TDC.

10.6 Bi-frequency Reference Oscillator

The bi-frequency mode reference oscillator was implemented with a structure similar to that used for the RF BAW DCO to get high tuning range and low-current consumption. Unlike the Pierce structure [26], this topology does not require a functional load capacitance to work properly (ideally $C_L > C_0$). A large C_L will impact significantly the power consumption and reduce the achievable tuning range. In the differential topology, the load capacitance can be reduced to the parasitic load of the MOS, the pads, and the switchable capacitor in its off state. This is made possible since above a given frequency that is controlled by the source to source decoupling capacitor placed below the cross-coupled pair, the circuit merely provides a negative conductance whatever the biasing level.

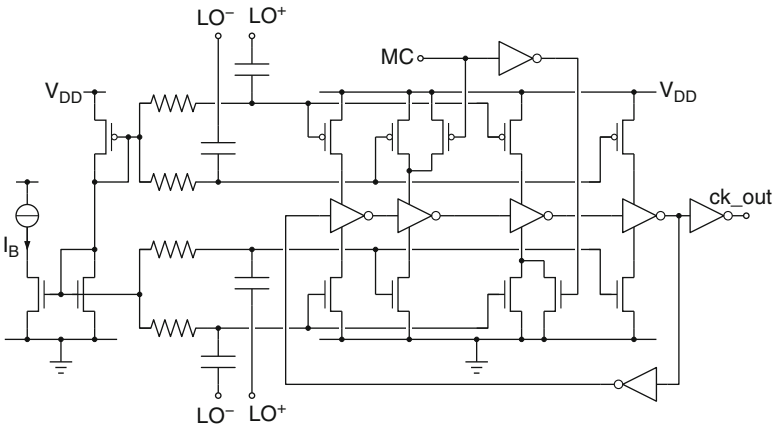

Fig. 10.5 Dynamic dual-modulus divider by 3–4 circuit with balanced input synchronizing signals. ©2009 IEEE. Reprinted, with permission, from [27]

Metal–metal finger-type capacitors switched by NMOS transistors were chosen to implement the bi-frequency mode since they yield high tuning capacitor with on/off ratios exceeding ten. The switched capacitor size is a trade-off between the tuning range and average current consumption of the oscillator which can both be evaluated with (10.2) and (10.6).

10.6.1 Divider Chains with Low-Power Dynamic Divider

The proposed architecture mandates for an efficient ultra-low-power divider implementation to maintain the overall consumption low, despite the four division chains that should run in parallel to generate the LO reference signal, the 16-MHz system clock, the 160–80-MHz IF signal, and the modulator clock used to obtain finer DCO frequency steps. A dynamic divider topology based on multi-modulus injection-locked ring oscillators with added control MOS transistors for synchronization or pulse swallowing is proposed to meet this challenge. All the divider cells are laid out with the size and pitch of standard digital cells using similar transistor sizes, yielding an extremely compact implementation and resulting in ultra-low-power and high-speed operation. Figure 10.5 shows the implementation of a differentially AC-driven dual-modulus divider by 3–4.

It consists of a ring of five inverter stages, among which four of them are current-starved, thanks to DC-biased series transistors to set the free-running oscillator speed. Opposite differential LO signals are then AC-coupled on the DC-biased transistors to alter the speed of each starved inverter, alternatively accelerating or decelerating the transitions in any two consecutive branches in an opposite manner.

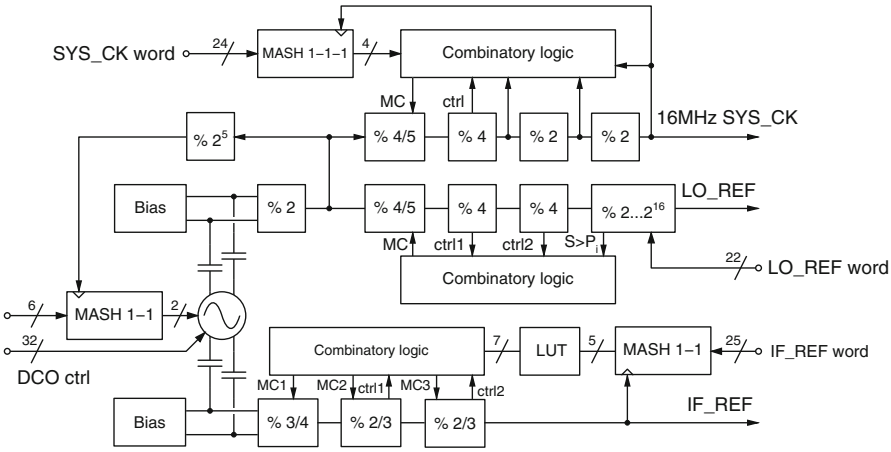

Fig. 10.6 Divider blocks arrangement. ©2009 IEEE. Reprinted, with permission, from [27]

Provided the free-running ring speed matches more or less that of the synchronizing signal and depending on the level of the latter, a single transition is allowed every half period of the LO, yielding injection locking and division by four. Division by three is merely obtained by swallowing two half periods out of a complete cycle, thanks to short circuiting transistors that waive synchronization on some positive and negative transitions. Single-ended driven ring dividers can be built similarly with an odd number of starved inverters whose synchronization terminals are all tied together to the rail-to-rail input signal that only allows a single transition in the ring as its polarity changes.

A block diagram showing the arrangement of the different divider blocks used to generate the desired signals is shown in Fig. 10.6. The IF signal generation chain made of a /3–4/2–3/2–3 arrangement combined to a 20-bits MASH 1–1 modulator allows generating any fractional division ratio between 13 and 35, hence signals between 178 and 66MHz with a 2.32-GHz input. The control of the three multi-modulus dividers to achieve the sequence of integer dividing ratio computed by the modulator is ensured with a combinatorial logic block that makes use of simple control signals derived from the second and third dividers. The 16-MHz system clock divider chain is built with a fixed pre-division by two followed by a /4–5/4/2/2 structure, which is combined to a 20-bits MASH 1-1-1 modulator and a combinatorial logic block to yield any fractional division ratio between 134 and 154. The reference LO divider chain is formed of a /2/4–5/4/4 dynamic structure followed by a 16-bits synchronous flip-flop-based counter implementing any integer division ratio between 2^{13} and 2^{23} in steps of two. Eventually, the 6-bits modulator controlling the DCO interpolator capacitor network is clocked at some 36 MHz after fixed division by 2^6 obtained with a cascade of 6 dividers by 2.

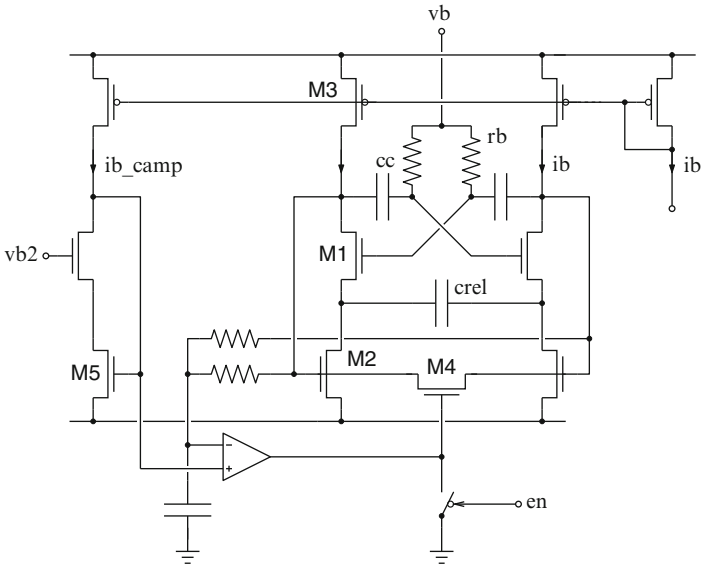

Fig. 10.7 Current-controlled relaxation oscillator with amplitude control implemented by adapting oscillator output impedance. ©2009 IEEE. Reprinted, with permission, from [27]

10.6.2 IF Relaxation Oscillator and Homodyne PLL

A quadrature relaxation oscillator has been designed to realize the wanted IF local oscillator [21]. The chosen basic topology is similar to the relaxation oscillator proposed in [28] and is depicted in Fig. 10.7.

Quadrature signals are obtained by duplicating the structure of Fig. 10.7 and using coupling transistors [21]. To reach the wanted higher frequencies with as low power as possible, the oscillator is kept in quasi-harmonic mode; the gates and drains of the main transistors M1 are DC decoupled by capacitors C_C, and the gates are biased via resistors R_b such that transistors M2 stays saturated. The gates of transistors M1 will be referred to as the oscillator output nodes. Analysis leads to the following relations:

$$g_{ms} \cdot R_{out} \geq 2 \left(\frac{1}{n} - \frac{1}{\alpha} \right)^{-1}, \tag{10.7}$$

$$\omega = \frac{g_{ms}}{2 \cdot C_{out}} \cdot \sqrt{\frac{1+\beta}{\alpha \cdot n} - \frac{1}{\alpha^2}}. \tag{10.8}$$

Self-sustained or growing oscillation amplitude is reached when relation (10.7) is verified, with R_{out} defined as real part of the parallel output impedance, n, the slope factor accounting for the body effect [29], and α is the ratio C_r/C_{out}. The source

transconductance g_{ms} is defined for transistors M1, and β is the ratio between the source transconductances of M2 and M1, $\beta = g_{ms,2}/g_{ms,1}$. The capacitance C_{out} merges all parasitic and load capacitance at the output nodes. Capacitance C_C is considered high enough to be neglected. The corresponding oscillation pulse frequency ω is described by relation (10.8). It can readily be derived from (10.7) that the capacitance ratio parameter α has to be greater than n. According to (10.8), the oscillation frequency can be controlled by varying the bias current I_C in order to modify the value of g_{ms}. If the value of R_{out} varies also proportionally to keep the condition (10.7) verified, the circuit of Fig. 10.7 implements a current-controlled oscillator (CCO). Meeting this last condition requires amplitude control.

1. *Amplitude control*: The wide variation of bias current needed to tune the oscillation frequency over the required range also creates wide variations in the oscillator operating point, more specifically in the steady-state oscillation amplitude [28]. In order to increase the tuning range, linearize the current-to-frequency characteristic, and allow the oscillator to stay in the quasi-harmonic mode over the entire range, an amplitude control has been designed. The amplitude control circuitry is detailed in Fig. 10.7. The adaptive biasing of transistor M4 allows to vary the impedance at the oscillator output nodes to maintain the gain $g_{ms} \cdot R_{out}$ constant, according to relation (10.7), even when the g_{ms} varies. When oscillations amplitude grows, the nonlinearity of the I–V characteristic of common-source transistor M2 decreases its average gate voltage [26]. This mechanism can be used to implement an amplitude control with a feedback loop which settles the average M2 gate voltage to a reference value generated by a replicated branch.
2. *Voltage-to-current converter*: In order for the current-controlled oscillator to be used in a conventional charge-pump PLL, a transconductance stage or voltage-to-current converter (VCC) is needed. The gain of this stage allows setting independently the equivalent voltage-to-frequency gain K_{VCO}. To keep K_{VCO} and the VCC gain reasonable as well as to insure the stage linearity, the tuning range is divided into multiple sub-bands. When used in a PLL, a finite-state machine algorithm automatically switches between sub-bands and resets the differential loop filter at each switching event. A sufficiently large hysteresis is set in the out-of-range events detection to ensure stability. The designed transconductance stage is a degenerated differential pair as detailed in Fig. 10.8.

   The differential tuning voltage V_{tune} allows to change the oscillator bias current I_C. The degeneration resistor R_d determines the transconductance gain G of the stage according to $G = 1/(n \cdot R_d)$. Several sub-bands are implemented by using current DACs. Replicated branches detect upper and lower out-of-range events by comparing currents on both sides, and sub-band selection algorithm increments or decrements the discrete value N depending on which event actually occurred.
3. *IF PLL*: In order to be able to filter out noise coming from the divided BAW oscillator, a simple homodyne PLL is implemented with a reverse gain of one (no division in feedback). The loop is using the relaxation oscillator described

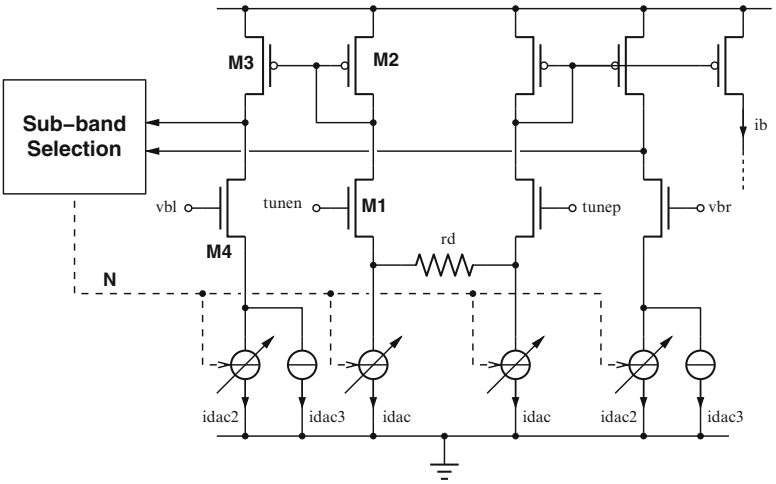

Fig. 10.8 Differential voltage-to-current converter with sub-band selection. ©2009 IEEE. Reprinted, with permission, from [27]

previously. The output frequency follows exactly the input frequency, with a filtering effect. The chosen implementation for the phase-frequency detector (PFD) is a classical tri-state PFD associated to a differential charge-pump circuit and currently external loop filter for greater flexibility.

10.7 Receiver

10.7.1 RF Front-End Using BAW Resonators

Using high-Q filters with good selectivity and low insertion loss in front of a receiver relaxes the overall linearity requirements and thus the power consumption. Moreover, heterodyne receivers need to filter out their image band in order to avoid signal degradation at the IF. Traditionally, off-chip SAW filters are used to this purpose. The feasibility of high-Q RF filters realized with BAW resonators has been demonstrated with comparable or even better results than SAW devices. An above-IC integration of a filter realized with thin film bulk acoustic wave resonators (FBAR) on top of a BiCMOS receiver front-end has been reported in [2], offering promising solutions for the integration of a complete RF SoC.

The BAW filter using SMR BAW resonator is placed in front of the LNA to eliminate out-of-band interferers and relax the linearity requirements. RF filters need correct impedance terminations in order to perform as specified. Input impedance matching is therefore a critical part of the LNA design. Classical LNAs as used in [2] require several critical inductors to provide a real input impedance

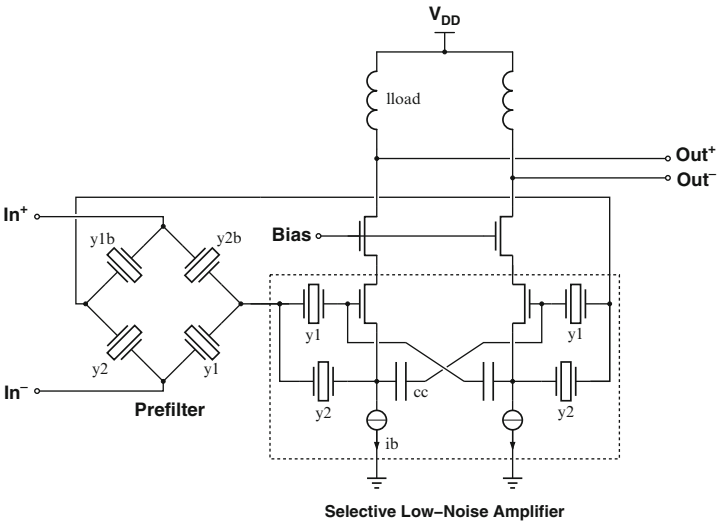

Fig. 10.9 Schematic of the selective LNA. ©2009 IEEE. Reprinted, with permission, from [27]

(typically 50-Ohm). On-chip inductors need a large silicon area and only offer limited quality factor. The main novelty of this approach is to co-design the LNA and the filter using BAW resonators. The use of only high-Q passive components helps in achieving better gain and noise figure performance.

The proposed selective LNA topology is shown in Fig. 10.9 [24]. A fully differential topology is chosen for better immunity to substrate and power supply noise. Moreover, balanced operation is mandatory for lattice filters and similar circuits such as the presented selective LNA to extend the available bandwidth to its maximum. The core of the LNA is analogous to the so-called gm-boosted LNA [30, 31] and presents a differential real impedance at the gates and sources of the transistors. Similar to lattice filters, the Y_1 and Y_2 resonators are appropriately detuned to achieve the widest bandwidth. These devices are mutually coupled through the network formed by the transistors and the cross-coupled capacitors C_C. In order to improve the out-of-band rejection, a prefilter stage is added in front of the circuit. This prefilter is a lattice network built from identical resonators than those used in the selective LNA. The main transistors are cascoded to improve reverse isolation and stability. The amplifier is loaded by an LC tank, the quality factor of which determines the output impedance and gain of the amplifier. The tank inductor is a differential integrated inductor using the five upper metal layers of the digital CMOS process. In order to realize a trimming capability on the load resonance frequency, several switched capacitors are added in parallel to the output nodes. A classical active switching mixer follows the LNA and downconverts the RF signal to the 160–80-MHz IF.

10.7.2 Intermediate Frequency and Baseband

A variable gain amplifier stage was added at intermediate frequency to provide additional gain for small input signals or attenuation for increasing the receiver linearity. Two further active switching mixers are used to convert the IF signal to baseband by mixing it with the quadrature IF LO signals. The baseband part of the receiver consists of an RC filter followed by a buffer, an active filter, and a limiters chain. Saturation levels of the limiter stages provide directly a logarithmic received-signal-strength indication (RSSI). The active filter consists of a third-order Butterworth low-pass Gm-C filter. Flexible channel filtering allows optimizing the bandwidth for several modulation indexes and bit rates.

10.8 Transmitter

10.8.1 Quasi-Direct Modulation with Fractional Divider

Direct modulation transmitter architectures, based on fractional PLLs, are advantageous to avoid LO pulling by the power amplifier, lead to better TX spectral content where spurs are avoided, and are very efficient from a power, hardware complexity, and area point of view. The noise shaping introduced by the $\Delta-\Sigma$ modulator limits however the achievable PLL bandwidth to a small fraction (typically 1/200) of the reference frequency, resulting in a limited data rate. Two-point modulation, requiring however frequent calibration of the varactor gain, alleviates the bandwidth/data rate trade-off at the expense of increased complexity. In the proposed superheterodyne architecture where all the tuning is performed at the IF frequency, the fractional division chain required for channel selection inherently allows high-speed modulation. A large gain in the achievable bandwidth is however obtained compared to classical PLLs since the frequency range lies very high between 80 and 160 MHz, and the level of the $\Delta-\Sigma$ modulator noise experiences no gain in the homodyne filtering PLL and up-converter mixer. A 1–2-Mb/s data rates compliant with *Bluetooth low-energy* [22] and *Zigbee* [23] standards are thus easily obtained without the need for any calibration.

10.8.2 DSB Up-Conversion Mixer with BAW Filtering

A double-sideband direct up-conversion mixer based on a pseudo-differential pair Gilbert mixer with current injection has been chosen to limit the number of stacked transistors and enhance low-voltage operation. Moreover, it can be demonstrated [32] that the injection of current in the pseudo-differential pair increases the conversion gain and the linearity, while decreasing the flicker noise coming from

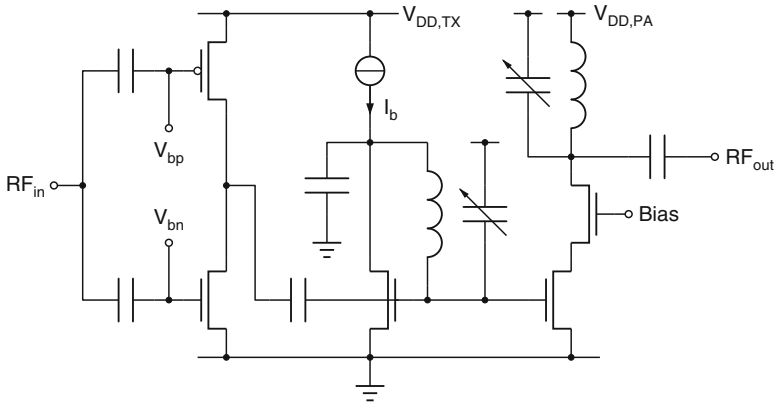

Fig. 10.10 Power amplifier simplified schematic (circuit is actually balanced. ©2009 IEEE. Reprinted, with permission, from [27]

the switching pairs. The availability of rail-to-rail IF signals allows driving the switching pairs directly. After proper biasing, the local oscillator signal is applied to the pseudo-differential pair. The active mixer output is loaded by an integrated differential inductance setting the mixer gain and input impedance of the following one-stage differential lattice BAW filter that is used to reject unwanted sidebands such as LO feedthrough, image, and IF harmonics. In order to present a well-defined output impedance to the BAW filter, another differential inductor is used to form a resonant load together with the input capacitance of the following amplifier stage. On-chip switched capacitors are present to adjust LC resonances. The filter input/output impedances have to be maximized to give the highest mixer gain and reduce the current consumption. However, the mean impedance of the constituting resonators determined by the shunt dielectric capacitance, C_O, has to be increased accordingly to maintain a rather flat response within the filter bandwidth, reducing the size, parallel Q factor, and achievable coupling of the resonators [33]. As a result, both the filter insertion loss and bandwidth are degraded when the resonator impedance is increased. To limit the power dissipation of the mixer, an impedance of 600 Ω was chosen for the resonators and the equivalent parallel resistance of the inductors defining the input/output impedance of the filter, although it was already considered somehow too high for the resonators.

10.8.3 CMOS Power Amplifier

The filtered signals coming from the up-converter mixer are amplified by an integrated CMOS power amplifier, depicted in Fig. 10.10. The preamplifier stage is a class B push–pull amplifier. The input signal is decoupled by means of integrated

capacitors while the biasing of the stage can be controlled by a programmable current. The structure is differential to reduce even order distortions. The differential power output stage is formed by cascoded transistors biased for class AB operation. The output stage is matched to 100 (differential), and another one-stage BAW filter may be used to further reduce unwanted sidebands. The LC resonance frequency at the interface can similarly be corrected by means of a switched capacitors bank to compensate possible detuning. Finally, on-chip decoupling of the power supplies has been used to filter out current peaks on the supply line. An additional advantage is brought by the differential structure, which allows to limit the magnitude of the supply AC current peak and to move the corresponding spectral component at twice the operating frequency [33], mitigating the on-chip decoupling requirements.

10.9 Experimental Results

Over the last years, several different circuits were realized and tested, ranging from simple RF blocks to the complete transceiver system shown in Fig. 10.2. The latest CMOS version of the transceiver chip is shown in Fig. 10.11. The results presented here have been chosen as the most representative of the combination of classical CMOS ICs and MEMS. All circuits have been integrated in compatible 1P6M 0.18-µm generic CMOS processes. The overall current consumption figures under 1.5-V supply voltage for the latest version of the transceiver chip are described in Table 10.1.

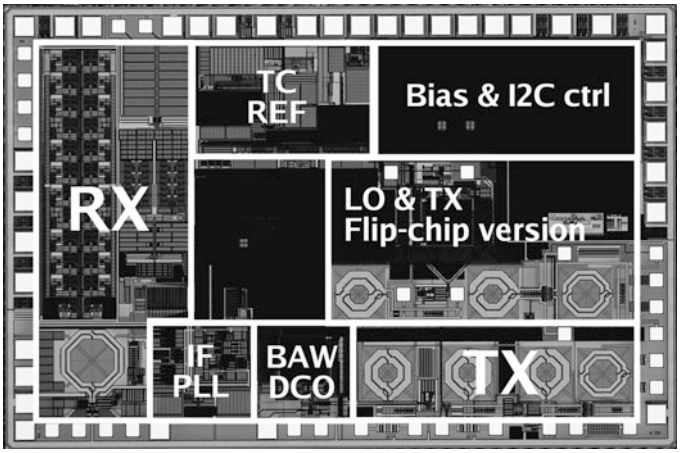

Fig. 10.11 Photograph of the integrated transceiver. ©2009 IEEE. Reprinted, with permission, from [27]

Table 10.1 Overall current consumption figures

Description	Current (mA)
BAW oscillator	0.9
GEN_IF divider chain	1.2
BAW ADPLL + CK 16-MHz dividers	1
BAW frequency tuning	0.2
IF CCO + PLL	0.65
Receiver	2.2
Up-converter	3.3
Preamplifier	2
Power amplifier	10.8
Total Rx mode	6.15
Total Tx mode @ $P_{out} = -4$ dBm	20.05

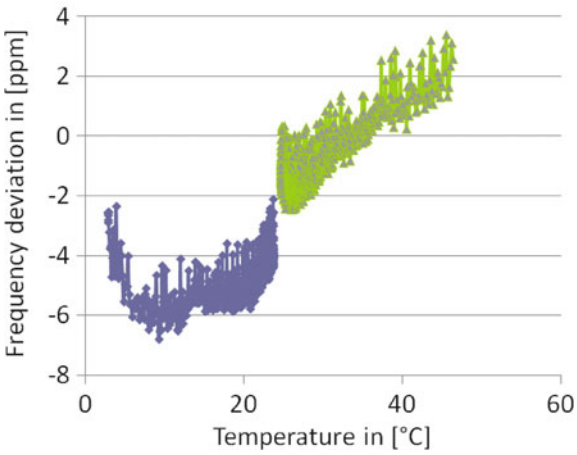

Fig. 10.12 Frequency stability of the 32'768-Hz temperature compensated clock. ©2009 IEEE. Reprinted, with permission, from [17]

10.9.1 32-kHz Reference Clock

Figure 10.12 shows a measurement of the compensated clock frequency deviation from 32'768 Hz versus temperature over a 0–50 °C range after individual calibration of the thermal sensor and resonator pair at three temperatures. The measurement was taken in two steps: after heating the sample at 50 °C followed by a temperature frequency pairs data-logging during its return to ambient and similarly after cooling down to 0 °C. The clock stability is better than 10 ppm and shows a remaining linear dependence of 0.25 ppm/°C, demonstrating a reduction by more than two orders of magnitude compared to the uncompensated resonator drift. It is currently limited by the accuracy of the calibration when determining the compensation parameters. The standard deviation of the frequency at a stable temperature is close to 0.6 ppm, in agreement with the product of the temperature sensor resolution and initial TCF of the resonator.

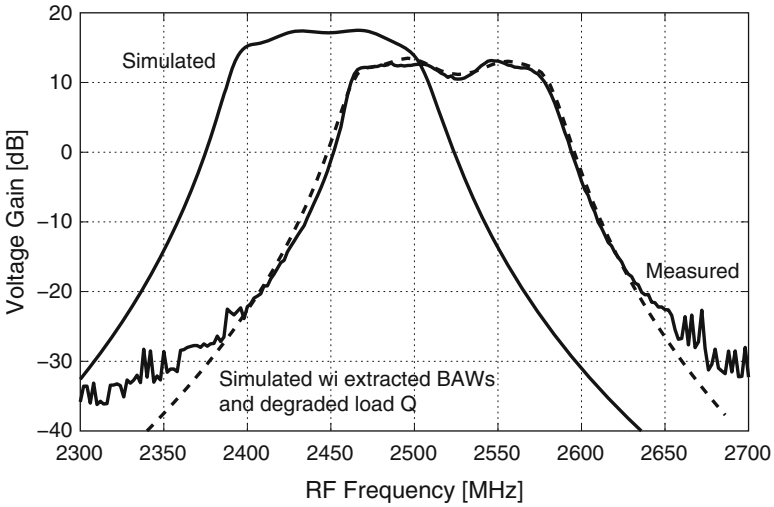

Fig. 10.13 Measured and simulated RF front-end voltage gain. ©2009 IEEE. Reprinted, with permission, from [27]

The average current consumption of the electromechanical oscillator amounts to 3 µA, and the divider chain draws 200 nA yielding multi-year autonomy with a coin-size battery. The temperature sensor consumes 30 µA but is duty cycled to yield negligible power overheads.

10.9.2 Receiver Front-End

This result highlights the selectivity attained by the first integration of the RF front-end [24]. To measure the circuit, both chips were bonded together and to external I/Os directly on a printed circuit board. In order to convert the single 50 generator output into a 100 differential input to the circuit, an external passive balun is used, resulting in a slight degradation of the overall noise figure due to its −1-dB insertion loss. The RF selectivity is measured with a constant IF frequency (90 MHz), and the results are shown in Fig. 10.13. It can be seen that measurements are close to the simulated results, with a center frequency shifted upward by 80 MHz. This is due to a thinner piezoelectric layer than expected in the resonators. Though this effect can be compensated by adjusting the thickness of an extra loading layer on the resonators, it requires an additional trimming step which was not performed for this prototype but is routinely achieved for industrial-level BAW resonators production [20]. The achieved image rejection is above 50 dB for the chosen IF. The overall voltage gain of 12 dB is 6 dB lower than simulation, and the front-end noise figure is degraded accordingly. This is mainly due to a lower Q than expected for the LC load.

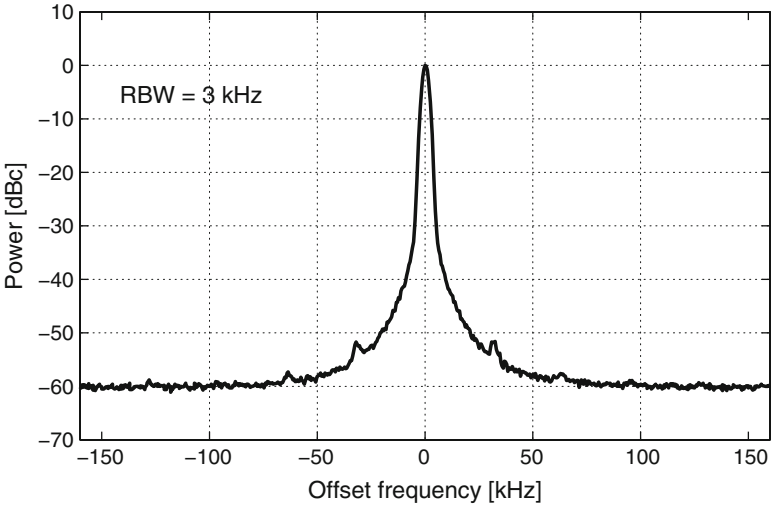

Fig. 10.14 Measured locked ADPLL spectrum (CF=2.3 GHz). ©2009 IEEE. Reprinted, with permission, from [27]

The measured noise figure for the overall front-end amounts to 11 dB. While this value may seem unduly high, it has to be noted that it includes all insertion losses in the balun and the filter, as well as the mixer noise in some proportion depending on the LNA gain. Linearity figures are a 1-dB compression point of −26 dBm and input IP3 of −16.1 dBm. The front-end consumes a total of 1.5 mA under 1.2 V.

10.9.3 Frequency Synthesis

Figure 10.14 shows the spectrum of the BAW DCO that is phase locked to a 32-kHz XTAL reference. A proportional, integral digital loop filter with a crossover frequency of 10 Hz was implemented on an FPGA to suppress the reference noise. The reference spurs are below −50 dBc and do not impact the system performances. Figure 10.15 shows relevant phase-noise plots of the synthesizer. The BAW DCO free-running spectrum (○ symbol), measured at the PA output when the IF port of the up-converter mixer is DC biased, is unexpectedly dominated by $1/f^3$ noise up to 1 MHz where it reaches −125 dBc/Hz, compared to a three decades lower noise corner reported previously with a DCO version with a slightly bigger area [21]. The reference IF signal obtained by fractional division from the BAW DCO (□ symbol) exhibits a better close-in noise due to frequency scaling, until a plateau is reached at −120 dBc/Hz probably due to the folding of modulator noise induced by non-linearity. Then the high-frequency shaped modulator noise increases above that level with a maximum at half the IF frequency. The same signal filtered by the homodyne PLL of the IF relaxation oscillator with a 1-MHz bandwidth (◇ symbol)

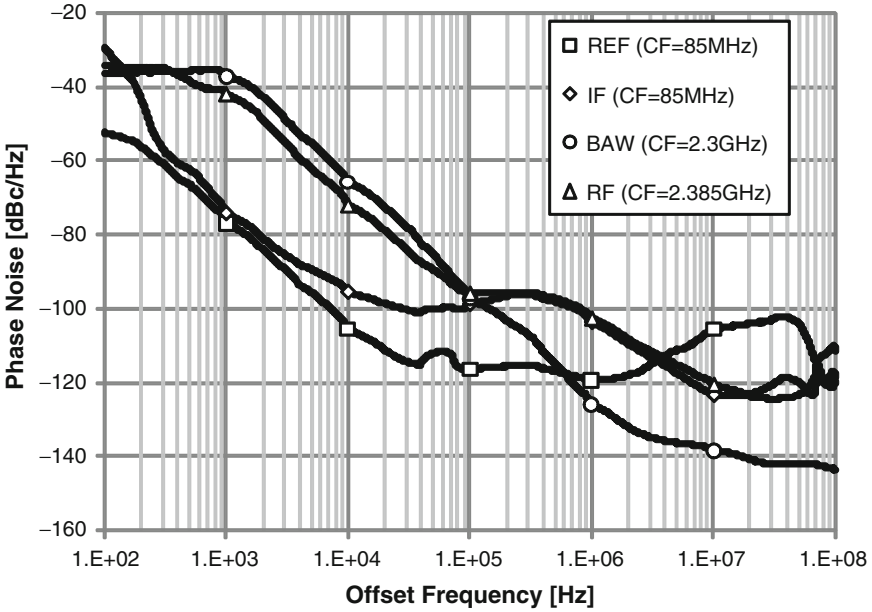

Fig. 10.15 Relevant phase noise plots. BAW LO; REF after BAW LO division; IF after PLL filtering, RF after BAW LO and IF mixing

inherits from the good close-in noise of its reference but suppresses its far-from-carrier noisy content. The IF in-band noise level reaches 95 dBc/Hz and is reduced to 113 dBc/Hz at 3-MHz offset. The unmodulated transmitter spectrum obtained after IF and RF LOs mixing merely consists of the sum of its two contributors and is hence not surprisingly determined by the low-Q IF oscillator phase noise at large offset (△ symbol). Gladly, the degradation is mitigated since a given phase noise level is easier to reach for lower frequencies.

10.9.4 Receiver Bit-Error Rate

Figure 10.16 shows a measurement of the receiver BER when a 100-kbit/s 2-FSK 23-bit long pseudo-noise sequence with a modulation index of eight is applied to the LNA input (without the front-end BAW filter), and the IF division chain is operated in fractional mode ($\Delta-\Sigma$ modulated). I & Q limited signals are fed to an external FPGA where the demodulator and clock recovery functions are implemented. A sensitivity of 87 dBm at a BER of 0.1% is obtained, corresponding to a degradation of 15 dB when compared to performances measured on a previous chip operated with external LOs. This discrepancy still requires further investigation to be fully explained, but 3 dB can be attributed to the noise coming from the image

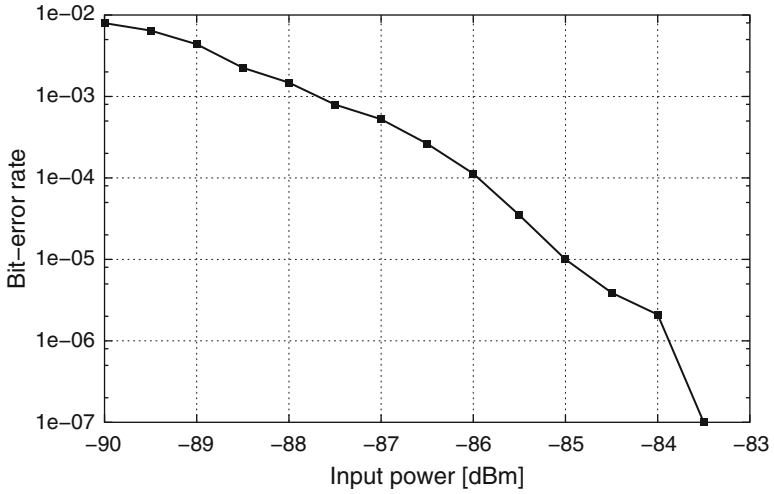

Fig. 10.16 Measured receiver bit-error rate at 100-kbps 2-level FSK. ©2009 IEEE. Reprinted, with permission, from [27]

band since the RX BAW filter was not mounted. Possible loss cause might be due to a non-optimal demodulator implementation, poorly matched LNA input impedance, and rather long printed circuit board RF tracks.

10.9.5 Modulated TX Signal

Figure 10.17 shows the non-modulated TX spectrum together with a 1-Mb/s GFSK-modulated signal that is generated on an FPGA and introduced in the fractional division chain used to produce the IF PLL reference frequency. The frequency deviation is set to 250 kHz, resulting in a 0.5 modulation index. Gaussian filter bandwidth symbol period product BT is 0.5. The average integrated RF power is −3.73 dBm with a current consumption of 3.3 , 2 , and 10.8 mA for the DSB up-converter mixer, preamplifier, and power amplifier, respectively. The transmitter measurements have been performed without the two designed BAW filters since the measurements of the 600-Ω intermediate filter stage have shown poor overall performance (2.5 dB in-band ripple, 3–4-dB insertion loss). The 100-Ω output stage filters have much better performances with an insertion loss of 11.5 dB and much reduced in-band ripple ($<$0.5 dB).

The association of the DSB mixer with the intermediate BAW filter is hence not attractive from a power consumption perspective. The additional current required to counterbalance the low impedance level limitation introduced by the BAW filter prevails over that spared by avoiding quadrature RF LO generation. If not compensated, this gain reduction further leads to reduced performances since the PA is operated in a lower efficiency class.

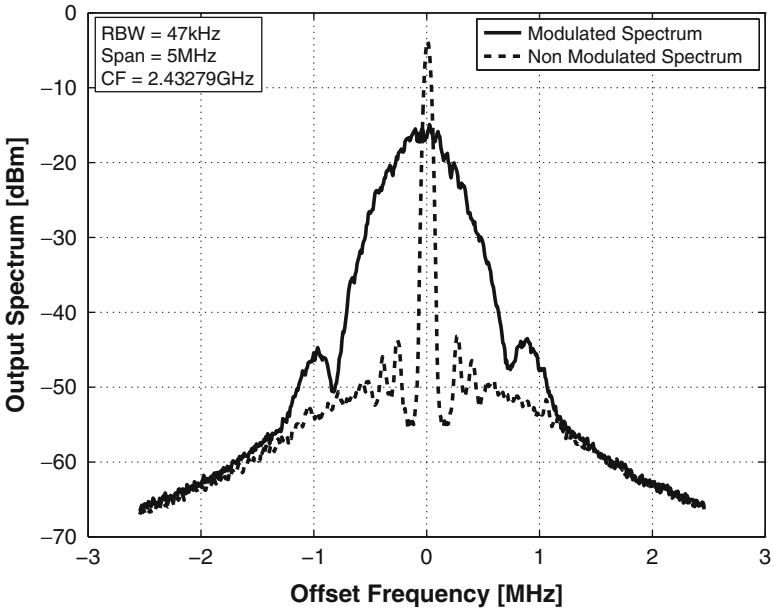

Fig. 10.17 Non-modulated RF carrier and modulated GFSK signal. ©2009 IEEE. Reprinted, with permission, from [27]

Figure 10.18 shows a picture of the chip-on board multi-chip module MEMS-based transceiver. The BAW resonator used in the DCO is visible to the right. To the left, an unpackaged 1-MHz silicon resonator was added for illustration purpose.

10.10 Conclusion

This chapter presents a new 2.4-GHz heterodyne radio architecture that is designed for an optimum use of MEMS devices such as high-Q BAW RF resonators and low-frequency Si resonators, paving the way toward an ultra miniaturized radio. The radio architecture takes advantage of the BAW high Q to obtain a low phase-noise fixed-frequency DCO. The limited tuning range of the BAW DCO is circumvented by shifting the tuning at the IF, hence relaxing the phase-noise requirement on the IF PLL, thanks to frequency scaling. A fast wake-up time is achieved, thanks to an ADPLL where the previous lock state can be memorized, allowing for the PLL to start almost in locked conditions despite its rather low loop bandwidth. The LO signal provided by the BAW DCO is divided by low-power dynamic fractional divider chains generating an IF frequency comprised between 66 and 178 MHz together with the 16-MHz system clock from a 2.32-GHz input. The IF PLL is built around a quadrature current-controlled relaxation oscillator with amplitude control.

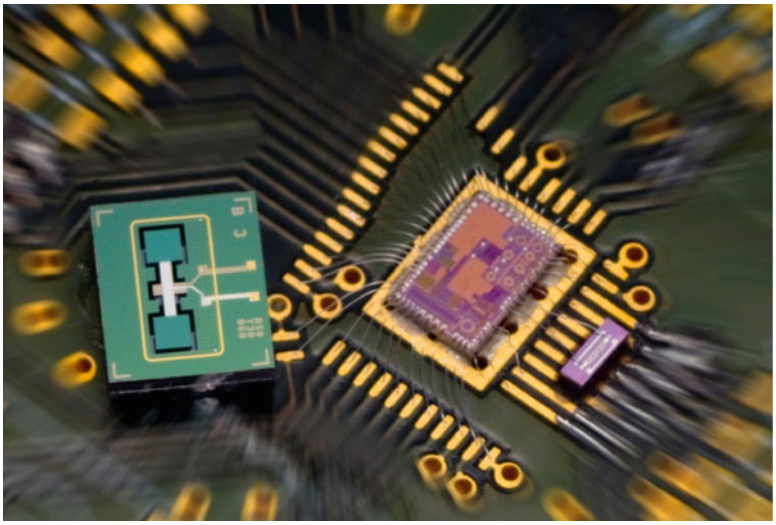

Fig. 10.18 Multi-chip module implementation showing Si Res, RF BAW resonator, and RFIC

BAW resonators are also used in the RF receiver front-end combining the functions of interferer and image rejection, impedance matching, and selective amplification, as well as in the transmit path for spurious components filtering. The radio circuit has been integrated in a 1P6M 0.18-μm generic CMOS process and assembled with BAW resonators. The complete radio has been validated demonstrating the advantage of this new architecture. The current consumption is 6.15 mA in receive mode with a sensitivity of −87 dBm at 100 kb/s and 20 mA in transmit mode with −4-dBm output power under 1.5 V. The frequency synthesizer has been measured using a 32-kHz external crystal. The free-running BAW DCO consumes about 900 μA and shows a −125-dBc/Hz phase noise at 1 MHz offset. The IF signal achieves close-in phase noise below −95 dBc/Hz and reaches −113 dBc/Hz at 3-MHz offset. The IF PLL consumes about 650 μA, whereas the fractional divider consumes about 1 mA. The receiver front-end achieves a voltage gain of about 12dB with a 1-dB compression point of −26 dBm and IP3 of 16.1 dBm, a noise figure of 11 dB for a 1.5-mA current consumption.

References

1. V. Peiris, C. Arm, S. Bories, S. Cserveny, F. Giroud, P. Graber, S. Gyger, E.L. Roux, T. Melly, M. Moser, O. Nys, F. Pengg, P.D. Pfister, N. Raemy, A. Ribordy, P.F. Ruedi, D. Ruffieux, L. Sumanen, S. Todeschini, P. Volet, in *IEEE International Solid-State Circuits Conference Digest of Technology Papers*, 2005, pp. 258–259
2. M.A. Dubois, J.F. Carpentier, P. Vincent, C. Billard, G. Parat, C. Muller, P. Ancey, P. Conti, IEEE J. Solid State Circ. **41**(1) (2006)

3. G.J. Carchon, R.W. De, E. Beyne, IEEE Trans. Microw. Theor. Tech. **52**(4), 1244 (2004)
4. A. El-Hoiydi, in *Proceedings of 7th International Symposium on Computers and Communications*, 2002, pp. 685–692
5. B.P. Otis, J.M. Rabaey, IEEE J. Solid State Circ. **38**(7), 1271 (2003)
6. B. Otis, Y. Chee, J. Rabaey, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2005, pp. 396–606
7. M. Aissi, E. Tournier, M. Dubois, C. Billard, H. Ziad, R. Plana, in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, 2006
8. M. Aissi, E. Tournier, M.A. Dubois, G. Parat, R. Plana, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2006, pp. 316–317
9. S. Rai, B. Otis, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2007, pp. 576–623
10. P. Vincent, J.B. David, I. Burciu, J. Prouve, C. Billard, C. Fuchs, G. Parat, E. Defoucaud, A. Reinhardt, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2008, pp. 478–479
11. W.T. Hsu, J.R. Clark, C. T.-C. Nguyen, in *International Electron Device Meeting Digest of Technical Papers*, 2000, pp. 399–402
12. W.T. Hsu, C.T.C. Nguyen, in *Proceedings IEEE 15th International Conference on Micro Electro Mechanical Systems*, 2002, pp. 731–734
13. R. Melamud, B. Kim, M.A. Hopcroft, S. Chandorkar, M. Agarwal, C. Jha, T. Kenny, in *Proceedings of IEEE 20th International Conference on Micro Electro Mechanical Systems*, 2007, pp. 199–202
14. J. Baborowski, C. Bourgeois, A. Pezous, C. Muller, M.A. Dubois, in *Proceedings of IEEE International Frequency Control Symposium*, 2007, pp. 1210–1213
15. P. Gupta, E. Radza, W. Chen, R. Sheridan, R. Sheridan, R. Melamud, M. Lutz, A. Partridge, K. Petersen, Leveraging Standard IC Packaging for MEMS Oscillators. http://www.sitime.com/news/articles/LeveragingstandardICpackaging.pdf. Accessed date 29th January 2010
16. K. Sundaresan, G.K. Ho, S. Pourkamali, F. Ayazi, IEEE J. Solid State Circ. **42**(6), 1425 (2007)
17. D. Ruffieux, A. Pezous, A.C. Pliska, F. Krummenacher, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2009, pp. 210–212
18. D. Ruffieux, J. Chabloz, C. Muller, F.X. Pengg, P. Tortori, A. Vouilloz, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2008, pp. 522–523
19. R. Staszewski, P. Balsara, *All-Digital Frequency Synthesizer in Deep-Submicron CMOS* (Wiley, New York, 2006)
20. E. Schmidhammer, H. Heinze, M. Schmiedgen, M. Mayer, A. Link, in *Proceedings IEEE Ultrasonics Symposium*, 2006, pp. 329–332
21. J. Chabloz, D. Ruffieux, A. Vouilloz, P. Tortori, F. Pengg, C. Muller, C. Enz, in *Proceedings of European Solid-State Circuits Conference ESSCIRC*, 2007, pp. 492–495
22. Bluetooth special interest group [Online]. http://www.bluetooth.org. Accessed date 29th January 2010
23. The Zigbee Alliance [Online]. http://www.zigbee.org. Accessed date 29th January 2010
24. J. Chabloz, C. Muller, F. Pengg, A. Pezous, C. Enz, M. Dubois, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2006, pp. 1244–1253
25. D. Ruffieux, in *Proc. European Solid-State Circuits Conference ESSCIRC*, 2002, pp. 85–88
26. E.A. Vittoz, M.G.R. Degrauwe, S. Bitz, IEEE J. Solid State Circ. **23**(3), 774 (1988)
27. D. Ruffieux, J. Chabloz, M. Contaldo, C. Muller, F.X. Pengg, P. Tortori, A. Vouilloz, P. Volet, C. Enz, IEEE J. Solid State Circ. **44**(1), 228 (2009). DOI 10.1109/JSSC.2008.2007447
28. B. Song, H. Kim, Y. Choi, W. Kim, in *AP-ASIC'99, The First IEEE Asia Pacific Conference on ASICs.*, 1999, pp. 154–157
29. M. Bucher, C. Lallement, C. Enz, F. Thodoloz, F. Krummenacher, The EPFL-EKV MOSFET model equations for simulation. Tech. rep., Ecole Polytechnique Fdrale de Lausanne (1998). http://legwww.epfl.ch/ekv
30. X. Li, S. Shekhar, D. Allstot, IEEE J. Solid State Circ. **40**(12), 2609 (2005)

31. D. Joos, M. Borremans. Differential Low Noise Amplifier with Low Power Consumption, 2007
32. T.J. Jung, K.S. Cheung, J. Lau, in *IEEE International Symposium on Circuits and Systems*, vol. 4, 2001, pp. 378–381
33. C. Muller, M.A. Dubois, in *Proc. IEEE Ultrasonics Symposium*, 2008, pp. 1552–1556

Chapter 11
A Digitally Controlled FBAR Frequency Reference

Hiroyuki Ito, Hasnain Lakdawala, and Ashoke Ravi

Abstract Crystal oscillators have been the primary frequency reference sources used in communication circuits. However, their size, frequency of operation, and cost are hard to scale. Push toward miniaturization and low cost demands smaller form factor and crystal-oscillator alternatives, and MEMS frequency references have been manufactured recently. In this chapter, a digitally controlled FBAR (freestanding bulk acoustic resonator)-based oscillator is presented as an alternative frequency reference.

11.1 Introduction

A quartz crystal resonator has been an essential device for modern electronics. Near 10 billions of crystal devices are sold every year, and their usages have continued to increase with growth in market size of electronic equipments, especially recent RF wireless applications.

A large portion of the RF circuit functionality has been integrated into a CMOS chips thanks to miniaturization of CMOS devices, resulting in reduction of board size and of the number of components in RF modules. However, the crystal oscillator, which has been the primary frequency source used in communication circuits, still remains as a discrete element, even in the latest RF wireless transceiver modules. No other device could attain Q-factor, frequency precision, and stability of AT-cut crystal resonators and could replace them for more than half a century.

H. Ito (✉)
Tokyo Institute of Technology, Yokohama, Japan
e-mail: itohiro@ieee.org

H. Lakdawala • A. Ravi
Intel Corporation, Hillsboro, OR, USA

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_11,

One of the significant issues in recent electronic equipments is that their sizes, frequencies of operation, and costs are not easily scaled simultaneously.

Some companies have recently started manufacturing silicon-MEMS oscillators with lower prices than crystal oscillators for replacing crystal devices (http://www.sitime.com, http://www.discera.com). Their package sizes are still larger than other discrete elements such as chip capacitors, and there is an opinion that silicon-MEMS oscillators will not dominate by 2012 due to tough specifications relating to phase noise and temperature compensation (http://www.wtc-consult.com/english/news/press-releases/28.03.08_mems-oscillators.html). However, silicon-MEMS crystal alternatives are gradually becoming smaller form factor and are steadily penetrating the market.

This chapter introduces a MEMS frequency reference using a FBAR (free-standing bulk acoustic resonator) for RF wireless communication applications. Application of the FBAR for a frequency reference has been proposed in the literature [1]. Some FBAR oscillators have achieved very low-phase-noise performances comparable to crystal oscillators [1, 2]. One of the challenges faced in the design of a FBAR frequency reference is the inaccuracy due to manufacturing variations and temperature changes. We show ideas for calibrating frequency fluctuation and propose a circuit architecture that is a digitally controlled FBAR-based oscillator (DCFO) with very high frequency resolution and enough tuning range to cover frequency variations. The resolution is achieved by a digital-to-analog converter (DAC) composed of three subranged capacitor banks, with fine resolution being obtained by 7-level sigma–delta capacitive DAC. We also address challenges due to spurious resonance in the oscillator output induced by interconnect paths in the package. The interconnect resonance effects are eliminated by a resonance suppression circuit, implemented with a tuned circuit at the source of the gain stage.

The organization of this chapter is as follows: Characteristics of crystal alternatives and the FBAR are discussed in Sect. 11.2, and a whole DCFO system and a frequency calibration method are introduced in Sect. 11.3. In Sect. 11.4, details of the DCFO are explained. Then, measurement results are shown in Sect. 11.5, and the chapter is summarized.

11.2 FBAR Characteristics as a Frequency Reference

Table 11.1 shows performances of different types of resonators. Bonding wire, chip (discrete), and on-chip inductors are also presented as references and can be used for composing *LC*-resonators. *Q* of such inductors, and consequently that of resonators built with these inductors, is much lower than that of crystal, ceramic, or MEMS resonators. It is not feasible to use such inductors in a primary frequency reference source in RF communication systems.

As might be expected, crystal resonators have the best *Q*, precision, and stability of resonance frequency. Limitation of crystal resonators are that their resonance frequency and size are very hard to scale. Resonance frequency and size of the

Table 11.1 A comparison of several different types of resonators and inductors

Type	Q	(Self) resonance frequency	Precision of resonance frequency	Stability of resonance frequency	Area
Crystal resonator (http://www.epsontoyocom.co.jp/, FA-128)	**100k**[a]	54 MHz	**20 ppm**	**20 ppm** (100 degrees)	2.0×1.6 mm
Ceramic resonator ([3], CSTCE_XK)	Lower than crystal	30 MHz	450 ppm	300 ppm (70 degrees)	3.2×1.3 mm
MEMS resonator (SiTime (http://www.sitime.com))	75 k	Several hundred MHz	100 ppm	100 ppm (100 degrees)	2.5×2.0 mm[b]
FBAR	$< 3,000$	$<$ **Several GHz**	$> 1,000$ ppm	$< 3,000$ ppm (100 degrees)	$<$ **0.5 $\times$ 0.5 mm**
Bonding wire	< 30	GHz range	Very bad		Larger than on-chip inductors
Chip inductor	< 40	GHz range	Around 5%		0.6×0.3 mm
On-chip inductor	< 15	GHz range	Worse than crystal and FBAR		**Several hundred μm square**

[a] Typical value of crystal resonators. Note that this value is not actual specification of this device
[b] Oscillator package size

Table 11.2 A phase noise comparison among crystal and FBAR oscillators

Reference	Process	Center frequency (Hz)	Phase noise at 10 kHz offset (dBc/Hz)
Crystal [4]	90nm Si CMOS	26M 2.0G	$-152-$ **114**[a]
[1]	0.18μm Si CMOS	1.9G	**−100**
[5]	0.13μm Si CMOS	2.2G	−85[b]
[7]	0.13μm Si CMOS	2.1G	−95[b]
[8]	0.13μm Si CMOS	1.9G	−98
[9]	0.35μm SiGe BiCMOS	5.5G	−93
[6]	0.25μm SiGe BiCMOS	2.1G	−94[b]
[2]	Si bipolar	2.0G	**−112**

[a] If the oscillator signal is ideally up-converted from 26 MHz to 2.0 GHz
[b] These values are read off from graphs in their papers

crystal resonator in Table 11.1 are higher and smaller class in crystal resonators, respectively. However, resonance frequency is only 54 MHz, and package size of 2.0×1.6 mm is larger than other discrete components that typically have size of 0.6×0.3 mm or 0.4×0.2 mm in modern RF front-end modules. Size reduction of crystal devices may progress in the future; however, small-size crystal resonators will be expensive due to increase of fabrication (packaging) costs. Ceramic resonators have been used for automotive applications and are superior in price, oscillation start-up time, and shock properties. However, their frequency and size are almost identical to crystal resonators. Table 11.1 also shows an example of a MEMS resonator (non-FBAR), and the resonator (http://www.sitime.com) has almost the same Q, precision, and stability of frequency as a crystal devices. It could be a substitute for crystal devices in some applications but seems still to be large in size at this time.

In terms of resonance frequency and area, FBARs have advantages over other resonators. Chip cost of the FBAR can be lower than that of the crystal resonator. One of the concerns is the Q of the FBAR, i.e., phase noise. Table 11.2 shows phase noise characteristics of crystal and FBAR oscillators. Phase noise of a well-designed crystal oscillator is around -150 dBc/Hz at 10 kHz offset [4], which corresponds to about -110 dBc/Hz at 2-GHz oscillation frequency. Most FBAR oscillators have achieved phase noise of better than -90 dBc/Hz, and oscillators in CMOS [1] have attained a very low phase noise of -100 dBc/Hz. One of the significant results is the -112 dBc/Hz phase noise of the FBAR oscillator of [2], the same level as crystal oscillators. This means FBAR oscillators could become low-phase-noise primary frequency reference sources in some RF communication systems when only phase noise is considered.

Other remaining challenges in the design of FBAR frequency reference sources are frequency fluctuations due to manufacturing variations and temperature changes. Our FBAR resonance frequency fluctuates by over 4,000 ppm and needs to be calibrated by using some techniques. Fortunately, we can achieve a frequency-controllable FBAR oscillator with over 4,000 ppm tuning range [5, 6]. The next section presents an idea of frequency calibration in RF communication systems.

11.3 The Whole System and Frequency Calibration Scheme

Before discussing an RF communication system with the FBAR and frequency calibration scheme, the brief target specification of the FBAR oscillator is shown in Table 11.3. Target phase noise is the crystal-oscillator level, and the frequency tuning range has to cover manufacturing and temperature variations. Frequency tuning resolution has to be better than 100 ppb because the frequency accuracy required for system such as Global System for Mobile Communications (GSM) is in the 100 ppb range. Even the best crystal oscillator cannot achieve this accuracy without feedback correction [4]. Thus, some communication standards (like GSM and WiMAX) allow calibration of the frequency source using beacon information from the transmitter. We can apply frequency control signals from a base station to frequency calibration of the FBAR oscillator as shown in Fig. 11.1. This method can be combined with factory calibration to overcome FBAR inaccuracies.

The frequency calibration scheme for crystal oscillators is also shown in literature [4], and we can use the same technique. The automatic frequency control (AFC) loop in the GSM network is explained here as an example. A mobile terminal needs to receive the frequency control burst (FCB) signal from the base station before it transmits signals. The RF transceiver demodulates and decodes the FCB signal using its own local oscillator; then, it maps decoded signals to the constellation plane and calculates the frequency offset. A digital baseband sends AFC digital signals to the reference clock oscillator and adjusts it until 100 ppb frequency accuracy is achieved.

Table 11.3 The brief specification of the FBAR oscillator

Phase noise at 10 kHz offset	Better than −110 dBc/Hz at GHz oscillation frequency
Frequency tuning range	Over 4,000 ppm (6.8 MHz for 1.7 GHz resonance frequency)
Frequency accuracy (tuning resolution)	Better than 100 ppb

Resonance frequency of our FBAR is around 1.7 GHz

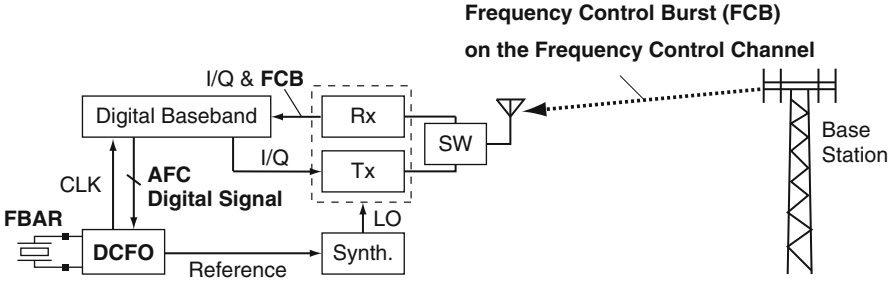

Fig. 11.1 An example of an RF communication system with the DCFO and a scheme of frequency calibration. *DCFO*: digitally controlled FBAR oscillator, *AFC*: automatic frequency control, *CLK*: clock, *Rx*: receiver, *Tx*: transmitter, *SW*: switch, *LO*: local signal, *Synth.*: synthesizer

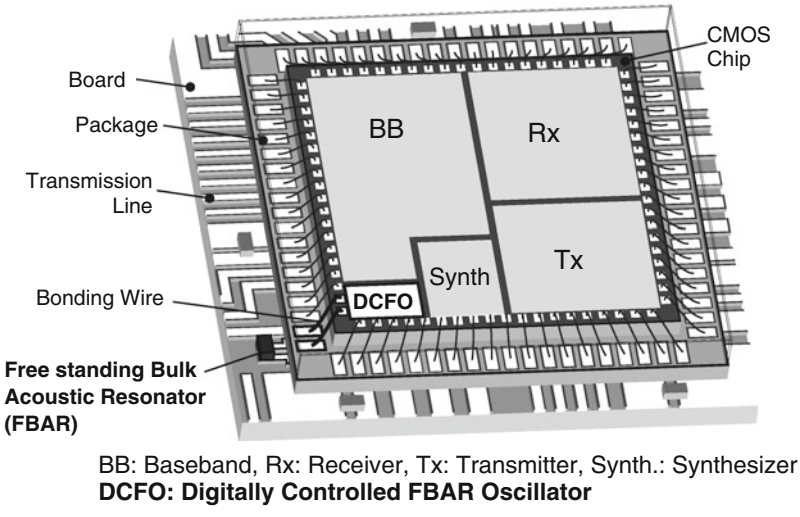

Fig. 11.2 An image of DCFO implementation

Required frequency resolution is extremely fine. Thus, we develop a digitally controlled FBAR oscillator (DCFO). Reasons are also the same as the crystal oscillator's case in [4] and are explained as follows:

1. Analog techniques for trimming require tiny varactors and very accurate voltage references. Tiny varactors have poor process-variation robustness, and accurate voltage references in recent sub-100-nm CMOS process have considerable design difficulties.
2. The AFC signal is a digital signal and has to be converted to the analog domain through a DAC if an analog FBAR oscillator (a voltage-controlled FBAR oscillator: VCFO) is applied. Design of the AFC DAC is also a difficult challenge because of the very low noise required for AFC DAC output signals that control varactors. Phase noise of the VCFO is easily degraded by any noise injected into varactor control lines. All tuning capacitors in the DCFO operate in low-gain saturation region or linear region and can achieve better noise robustness.
3. Ratio of maximum and minimum varactor capacitance is limited especially in sub-100-nm CMOS processes with low supply voltage, which reduces the frequency tuning range of the VCFO. Maximum to minimum capacitance ratio in the digitally controlled oscillator is limited only by the parasitic capacitance and is usually much larger than what an analog varactor can achieve [4].

The DCFO will be implemented into RF modules like Fig. 11.2. Considering an initial product, the FBAR will be placed on a printed circuit board. The fully integrated RF transceiver chip is packaged and is connected to the FBAR through bonding wires and transmission lines. The interconnect path including bonding wires and transmission line causes a stability issue of the DCFO, and a solution

is proposed in Sect. 11.4.2. Above-IC integration of the FBAR resonator would eliminate this problem. However, the additional fabrication process for the FBAR on CMOS wafers and three-dimensional integration of the FBAR still have cost and reliability issues.

11.4 A Digitally Controlled FBAR Oscillator

Figure 11.3 shows a block diagram of the proposed DCFO which consists of a core oscillator, the FBAR, a DAC including a second-order sigma–delta modulator, a phase controller, a programmable divider, and a buffer. Low-pass filters and buffer amplifiers (CMOS inverters) are inserted at coarse and moderate control inputs reducing noise coupled from digital circuitry. The clock of the sigma–delta modulator is supplied from the oscillator output through the divider and the phase controller which adjusts switching timing of the fine capacitors for minimizing phase noise. The core of the DCFO is explained next. We also show ideas for achieving very fine frequency resolution, an issue of bonding wires connecting the FBAR with the chip and solutions for it.

11.4.1 Overall Design of the Oscillator Core

First, circuit topology and transistor sizes in sub-100-nm CMOS process have to be carefully designed for achieving low phase noise such as less than -100 dBc/Hz at 10 kHz offset. Especially, flicker noise of MOS transistors in nanometer CMOS processes is an important concern.

The schematic of the oscillator core is shown in Fig. 11.4, and FBAR resonators can be modeled using the modified Butterworth Van Dyke model (mBVD) [10]. The basic topology is a differential oscillator introduced by [11]. One of the merits

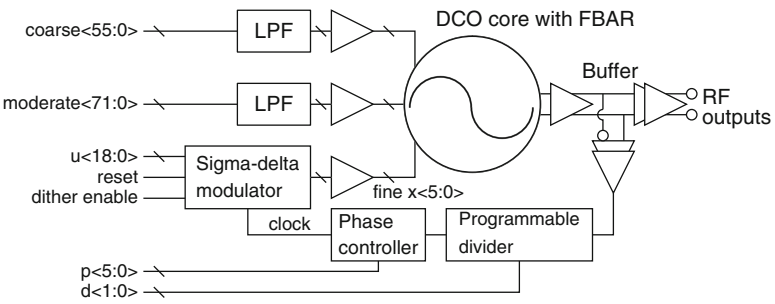

Fig. 11.3 Block diagram of the oscillator with the subranged capacitive DAC including a sigma–delta modulator, a phase controller, and a programmable divider

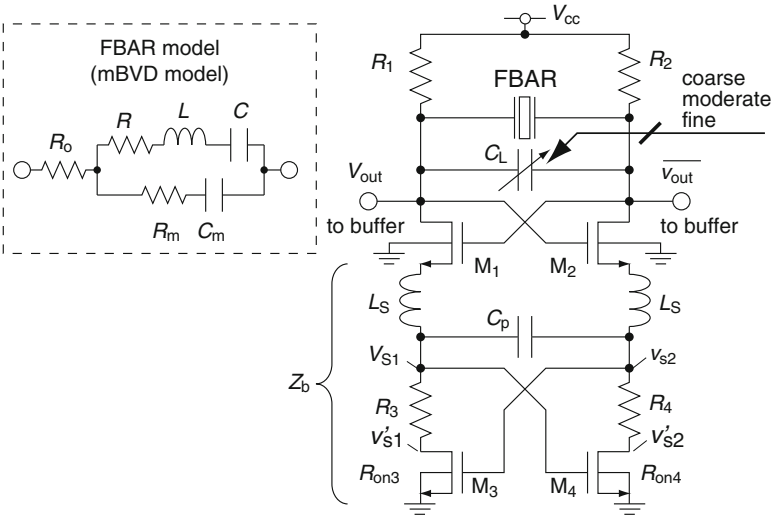

Fig. 11.4 Schematic of the DCO core and a FBAR model. ©2008 IEEE. Reprinted, with permission, from [13]

of this topology is that the number of transistors operating in saturation region, i.e., flicker noise sources, can be limited to the cross-coupled pair M1 and M2. A current source, which is generally applied for *LC*-type voltage-controlled oscillators (VCOs), does not seem to be suitable for very low-phase-noise oscillators because it also generates noises. We also need to care about the up-conversion of noise from the current source [12]. However, this topology is not immune to noise from the power supply. A voltage regulator may be required for implementing the DCFO in RF wireless systems.

Sizes of M1 and M2 are designed to minimize phase noise at lower offset frequencies. Flicker noise seems to be getting worse with transistor size reduction in advanced CMOS processes, and the phase noise of oscillators tends to have a higher corner frequency. Average power of flicker noises in a bandwidth of 1 Hz can be approximately modeled as

$$\bar{v_{\mathrm{n}}^2} = \frac{K}{C_{\mathrm{ox}} W L} \cdot \frac{1}{f}, \tag{11.1}$$

where C_{ox} is gate capacitance per unit area [14]. K is a process-dependent constant on the order of 10^{-25} $\mathrm{V^2F}$. Thus, one method in circuit design for reducing flicker noise of MOS transistors is to increase the product of gate width W and gate length L. When the minimum gate length L_{min} of process is used, large W is needed for reducing flicker noise. However, increase of W improves transconductance g_{m}, which degrades the thermal noise

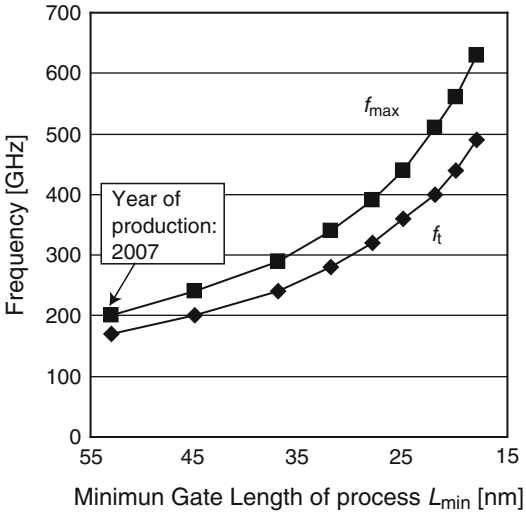

Fig. 11.5 Dependences of f_t and f_{max} on technology predicted by the International Technology Roadmap for Semiconductors (ITRS) [15]

$$\bar{i_n^2} = 4kT\gamma g_m,^1 \tag{11.2}$$

where $k = 1.38 \times 10^{23}$ J/K and γ are the Boltzmann constant and the coefficient determined by channel length or the drain-source voltage, respectively [14].

Thus, we should use larger gate length L. Then, the maximum transit frequency f_t and maximum frequency of oscillation f_{max} of large L transistors become of concern. f_t and f_{max} can be determined by

$$f_t = \frac{g_m}{2\pi(C_{gs} + C_{gd})} \tag{11.3}$$

$$f_{max} = \sqrt{\frac{f_t}{8\pi R_g C_{gd}}}, \tag{11.4}$$

and large-L devices obviously have lower f_t and f_{max}. C_{gs}, C_{gd} and R_g, are gate-source capacitance, gate-drain capacitance, and gate resistance, respectively. Figure 11.5 shows f_t and f_{max} dependence on the minimum gate length of process L_{min} predicted by the International Technology Roadmap for Semiconductors (ITRS) [15]. As of production year 2007, peak f_t and f_{max} are over 170 GHz. Since f_T and f_{max} of recent and future sub-100-nm CMOS processes are more than high enough to realize several-GHz-band analog circuits, increase of gate length L is an expedient way to improve noise characteristics of transistors and RF circuits. Obviously, it is to be noted that extremely large-L transistors consume large area

[1]The actual equation is $\bar{i_n^2} = 4kT\gamma g_{ds}$, where g_{ds} is the drain-source conductance with $V_{ds} = 0$ [14]. Equation (11.2) is for long-channel devices in a saturation region; however, we use it for simple discussion.

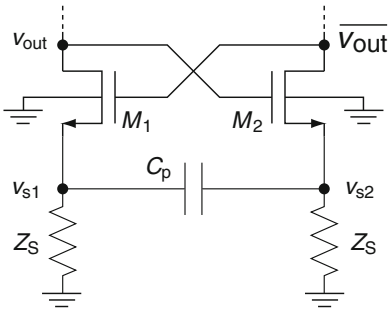

Fig. 11.6 The simplified bottom part of the DCFO at lower frequencies

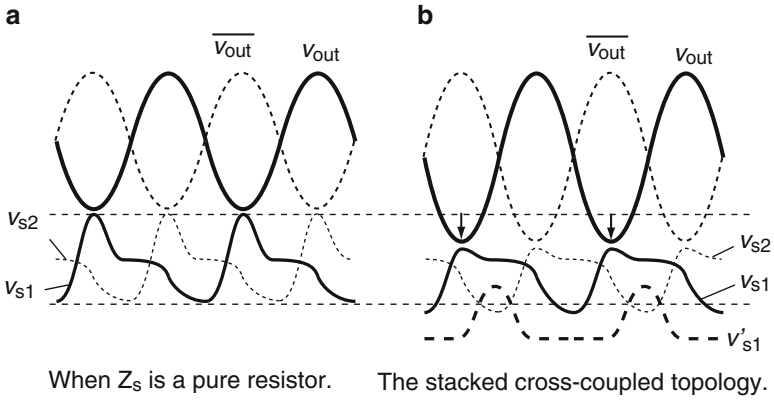

Fig. 11.7 Time-domain wave forms. (**a**) When Z_s is a pure resistor. (**b**) The stacked cross-coupled topology

and reduce frequency tuning range of the DCFO due to large gate capacitance. In our design, gate lengths of M1 and M2 are three times the minimum gate length of the 90-nm CMOS process.

Load resistors R_1 and R_2 are designed so that amplitude of oscillation signal is maximized. C_p, R_3, R_4, M3 and M4 contribute stability of oscillation. L_s is necessary to solve an implementation issue, and we will explain it in the next section. According to the literature [11], C_p has to be less than $n(C_L + C_m)$, where n is the transistor body effect factor. Small C_p avoids relaxation oscillation at lower frequencies but reduces oscillation amplitude, so the trade-off has to be considered in design.

R_3, R_4, M3, and M4 correspond to Z_s as shown in Fig. 11.6. Z_s reduces DC gain and is also necessary to stabilize the DCFO. However, Z_s is conventionally a pure resistor, which limits the lower level of signals v_{out} and $v_{\bar{out}}$ as shown in Fig. 11.7a. Thus, we used the stacked cross-coupled topology. The bottom cross-coupled pair (M3, M4), which mostly operates in the linear region and hardly generates flicker noise, slightly emphasizes signal amplitude as compared to the case of the pure resistor. When oscillation starts, $R_3 + R_{on3}$ is almost constant and stabilizes the DCFO at lower frequency, where R_{on3} is the source-drain resistance of M3. In the

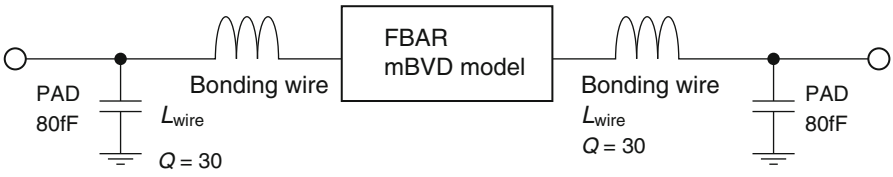

Fig. 11.8 A simplified model of the FBAR element along with the interconnect effects seen by the oscillators. ©2008 IEEE. Reprinted, with permission, from [13]

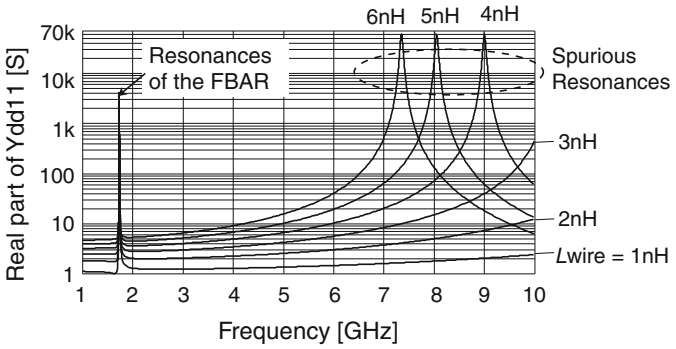

Fig. 11.9 Input admittance of the circuit of Fig. 11.8. ©2008 IEEE. Reprinted, with permission, from [13]

steady state, voltage swings of v_{s1} and v_{s2} change source-drain impedances of M3 and M4 (R_{on3} and R_{on4}), i.e., Z_s. When v_{s2} is high, the drain-voltage of M3 v'_{s1} decreases due to reduction of R_{on3}. This suppresses peak voltage of v_{s1} as shown in Fig. 11.7b. Then, the lower level of v_{out} is slightly emphasized. In simulations, oscillation amplitude increase of 44 mV and phase noise improvement of 0.9 dB/Hz at 10 kHz offset without power penalties are seen.

11.4.2 A Spurious-Resonance Suppressor

A CMOS chip and the FBAR are connected to each other through an interconnect path that includes bonding wires and transmission lines as shown in Fig. 11.2. Figure 11.8 is a simplified model of the FBAR as seen by the oscillator. The simulation results in Fig. 11.9 illustrate the effect of varying the length of the bonding wires, changing their inductance (L_{wire}). The differential input admittance of the FBAR with wires and pads shows the desired FBAR resonance around 1.7 GHz. A resonance at over 7 GHz is spurious and is induced by bonding wires, and the spurious-resonance frequency decreases as L_{wire} increases. The significant implementation issue is that the DCO core oscillates at the spurious frequency if L_{wire} is as large as 3 nH, which is realistic when the CMOS chip and the FBAR are placed on a board.

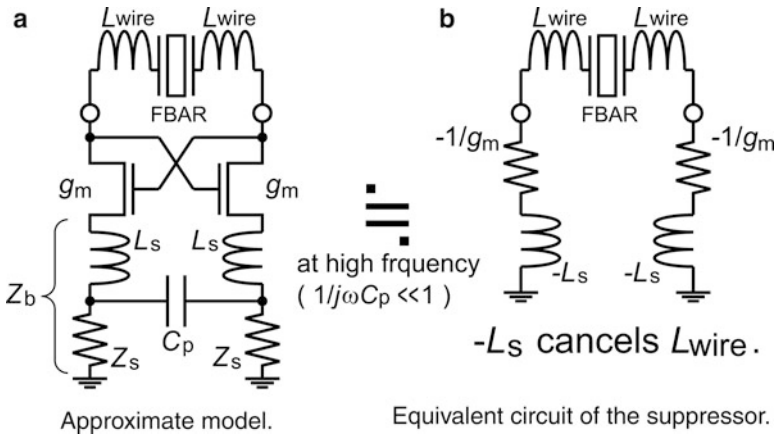

Fig. 11.10 The proposed spurious-resonance suppressor circuit. ©2008 IEEE. Reprinted, with permission, from [13]. (**a**) Approximate model. (**b**) Equivalent circuit of the suppressor

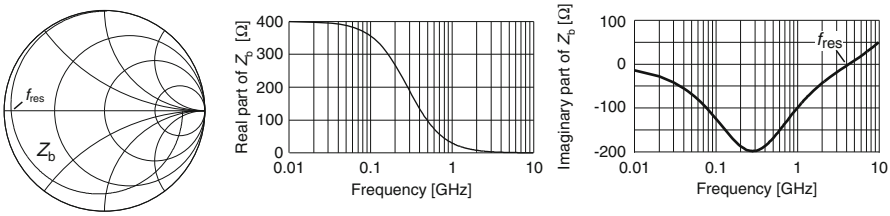

Fig. 11.11 Frequency dependence of Z_b. $L_s = 1.0$ nH, $Z_s = 400\ \Omega$, and $C_p = 0.7$ pF

One solution is to neutralize the effect of the parasitic inductor by exploiting a negative impedance circuit. The concept of the cancelation scheme is shown in Fig. 11.10. Coarse, moderate, and fine capacitors are excluded for simplicity. Odd-mode impedance of the DCFO bottom-part Z_b is

$$Z_b = \frac{Z_s + j\left\{4\omega^3 Z_s^2 L_s C_p^2 + \omega\left(L_s - 2Z_s^2 C_p\right)\right\}}{4\omega^2 Z_s^2 C_p^2 + 1}, \tag{11.5}$$

where ω is angular frequency. When frequency f is

$$f_{res} = \frac{1}{2\pi}\sqrt{\frac{2Z_s^2 C_p - L_s}{4Z_s^2 L_s C_p^2}}, \tag{11.6}$$

imaginary part of Z_b becomes zero.

Figure 11.11 shows frequency dependence of Z_b. f_{res} will be set at around the desired oscillation frequency f_{osc} for utilizing both effects of "Z_s and C_p" and "L_s". Because influences of Z_s and C_p should be dominant as frequency is lower than

f_{osc}, they can prevent the oscillator from latching and relaxation oscillation at low frequencies as described in the previous section. Imaginary part of Z_b increases at high frequency, then Z_b almost behaves as L_s at over f_{res}. Thus, when cross-coupled transistors operate in saturation region, an equivalent circuit model of a bottom portion at high frequencies can be represented by negative resistors $-1/g_m$ and negative inductors $-L_s$ as shown in Fig. 11.10b. Then, negative inductors $-L_s$ can cancel out parasitic inductors of bonding wires L_{wire} and can suppress spurious resonance.

A filter connected in parallel with the FBAR would also be able to suppress spurious resonance induced by bonding wires. However, this could degrade the oscillation signal amplitude and could introduce another spurious resonance caused by the filter. An advantage of this technique is that the spurious suppressor hardly affects desired oscillation characteristics and minimally degrades phase noise. Parasitic series resistors of inductors slightly degenerate the gain of cross-coupled transistors M1 and M2 and have to be made up by increasing the g_m of the cross-coupled pair.

11.4.3 Principle of Frequency Tuning

Three types of capacitors, coarse $C_c < 55:0>$, moderate $C_{md} < 71:0>$, and fine $C_f < 5:0>$, are connected in parallel to the FBAR as shown in Fig. 11.4. When the negative resistance is large enough, oscillation can be achieved at frequency of [11]

$$f_{osc} = \frac{1}{2\pi\sqrt{LC}}\left(1 + \frac{C}{2(C_m + C_L)}\right). \tag{11.7}$$

$C_L = C_c < 55:0> + C_{md} < 71:0> + C_f < 5:0>$ is total capacitance of capacitor banks, and oscillation frequency can be controlled by changing the number of capacitors through frequency control signals from the digital baseband as described in Sect. 11.3.

The frequency control scheme with three kinds of capacitors is very similar to the digitally controlled *LC*-based oscillator [16]. Coarse capacitors roughly calibrate oscillation-frequency variations caused by process and temperature variations. Moderate capacitors mainly compensate temperature fluctuation, and fine capacitors improve the frequency resolution. Details of these capacitors are explained in the following sections.

Note that Q of coarse, moderate, and fine capacitors are not important for this oscillator. Oscillation characteristics of this circuit are mostly determined by series resonance of the oscillator tank. Series Q of the FBAR is very high, on the order of 2,500, and most of the current flows into the FBAR at series resonance frequency due to very small impedance of the FBAR. Thus, series Q of the oscillator tank is almost determined by series Q of the FBAR and is hardly degraded by these capacitors.

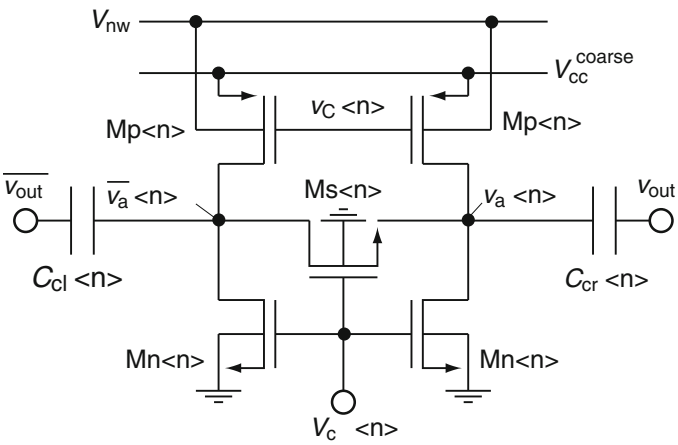

Fig. 11.12 Coarse capacitors

11.4.3.1 Coarse Capacitors

The schematic of a coarse capacitor is shown in Fig. 11.12. The n-th coarse capacitor consists of metal-insulator-metal (MIM) interdigital capacitors ($C_{cl} < n >$, $C_{cr} < n >$), an NMOS switch (Ms $< n >$), pull-up transistors (Mp $< n >$), and pull-down transistors (Mn $< n >$). Voltages $V_c < n >$ control states of coarse capacitors. This topology ensures suitable DC bias voltages for Ms $< n >$ at both on- and off-states. Maximum and minimum capacitance of the whole coarse bank is 770 fF and 270 fF in our design.

When the control voltage $V_c < n >$ is high, the NMOS switch Ms $< n >$ and pull-down transistors Mn $< n >$ turn on. Then, the n-th coarse capacitor is active. Pull-down transistors Mn $< n >$ set source and drain voltages of the NMOS switch Ms $< n >$ (DC voltages of $v_a < n >$ and $\bar{v_a} < n >$) to ground. This maximizes overdrive voltage of Ms $< n >$, which provides minimum on-resistance of the switch. The size of Mn $< n >$ can be minimized because pull-down transistors Mn $< n >$ supply only a DC level.

The coarse capacitor turns off when the control voltage $V_c < n >$ goes low. DC voltages of $v_a < n >$ and $\bar{v_a} < n >$ are V_{cc}^{coarse}, which enables high off-resistance of Ms $< n >$. Note that V_{cc}^{coarse} should be around half of the back-gate voltage V_{nw} of Mp $< n >$, i.e., about half of supply voltage V_{cc} of the oscillator. RF signals ($v_{sig} \simeq v_{out}$) swing around the DC level V_{cc}^{coarse} at drain and source nodes of Ms $< n >$. If $v_a < n >$ and $\bar{v_a} < n >$ become much higher than V_{nw}, forward bias voltage is applied to a diode at a junction of n-well and p-implantation layer of drain as shown in Fig. 11.13, resulting in distortion of RF signal v_{sig}.

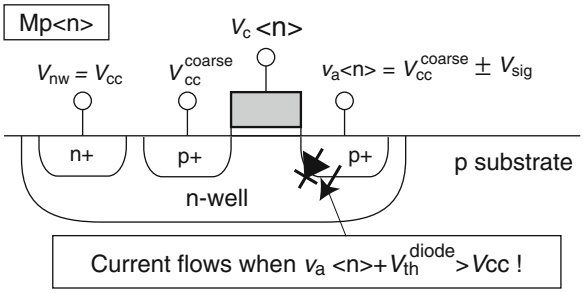

Fig. 11.13 Cross section of pull-up devices Mp < n > and bias conditions

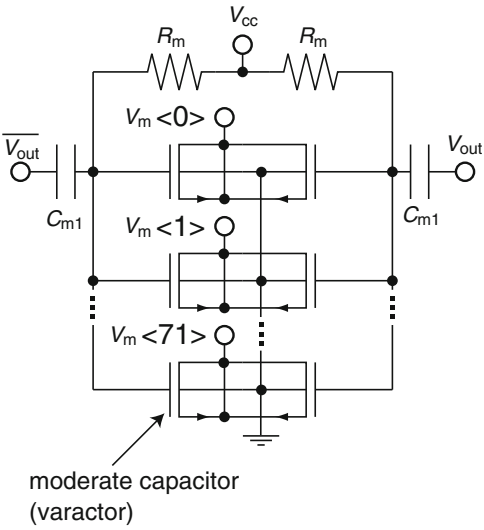

Fig. 11.14 A moderate capacitor bank

11.4.3.2 Moderate Capacitors

A moderate capacitor bank consists of NMOS varactors, MIM capacitors C_{m1}, and resistors R_m for DC bias as shown in Fig. 11.14. Nodes $V_m < n >$ control the number of active moderate capacitors. The moderate bank is designed to have enough frequency overlap between coarse and moderate tuning for ensuring margin of process variations. Maximum capacitance of the moderate bank is 80 fF in our design, which means capacitance of 1.4 fF is required for each moderate capacitor. A small-size varactor with such tiny capacitance is critically affected by process variations. Thus, the moderate bank includes capacitors C_{m1} in series with varactors, and effective capacitance per moderate capacitor can be reduced even though varactors are not tiny. R_m should be designed so that parasitic capacitance of R_m becomes small. High R_m improves Q of the moderate bank; however, high-value resistors have large parasitic capacitance and degrade tuning range. Q of the moderate bank is not important because it hardly affects the series Q of the resonator tank as described above.

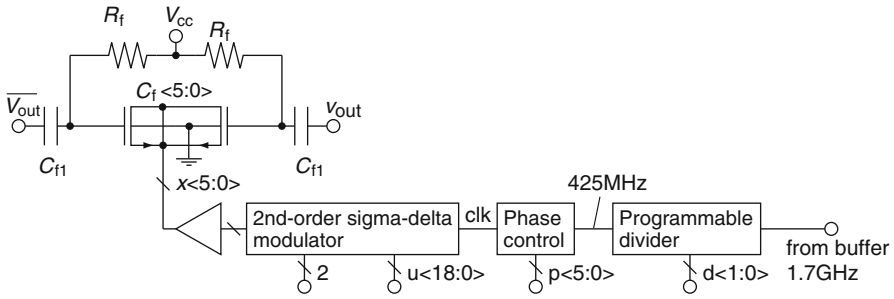

Fig. 11.15 A fine capacitance tuning system

11.4.4 Fine Frequency Tuning by a Sigma–Delta Capacitive DAC

Figure 11.15 shows the schematic of a fine bank with six NMOS varactors $C_f < 5:0>$, MIM capacitors C_{f1}, resistors R_f for DC bias, a second-order sigma–delta modulator, a phase controller, and a programmable divider. Sizes of NMOS varactors and resistors are the same as those in the moderate bank. MIM capacitors C_{f1} are designed to have 50% tuning overlap between the whole fine bank and each moderate capacitor.

The clock for the sigma–delta modulator is supplied through the programmable divider and the phase controller. The frequency division ratio of the programmable divider is set to 4. The phase controller changes timing of capacitance change that affects jitter of phase noise. The varactor capacitance should be changed at times when capacitors are fully discharged [16].

The second-order sigma–delta has a 19 bit input $u < 18:0>$ and drives 6 output-capacitors $C_f < 5:0>$ of the fine bank. The sigma–delta is an error feedback structure that breaks cascades of adders with register insertion to operate at the high clock frequency that is 1/4 the oscillator frequency (425 MHz) at very low power. Second-order sigma–delta modulators have better noise-shaping characteristics but are less stable than first-order sigma–delta [17]. The stability issue can be avoided by limiting the input signal pattern. A cycle of the frequency control signal from the digital baseband is much longer than a clock period of the sigma–delta, so an input signal (a control signal) of the sigma–delta $u < 18:0>$ can be roughly treated as DC. The stability of second-order sigma–delta modulators with DC input $|u|$ less than one in magnitude has been rigorously established [17]. If the full-scale input range is desired, combination of the first-order sigma–delta and pseudorandom bit sequence dithering would be helpful for avoiding stability and idle tone problems [4].

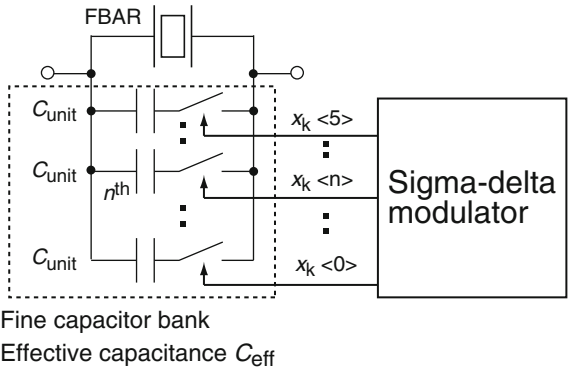

Fig. 11.16 Simplified schematic of the fine capacitor bank connected to the FBAR. ©2008 IEEE. Reprinted, with permission, from [13]

11.4.4.1 A Principle of Frequency Tuning

Each output bit of the sigma–delta $x_k < n >$ (= 0 or 1) at a cycle k turns on or off the n-th capacitor with value C_{unit}. The quantization noise from the sigma–delta is filtered by the bandpass response of the FBAR, which allows the capacitive DAC to have a high resolution. The effective capacitance of the fine bank over K cycles of the sigma–delta loop clock is

$$C_{\text{eff}} = \frac{\sum_{k=0}^{k=K-1} \sum_{n=0}^{n=N-1} x_k < n >}{K} C_{\text{unit}}. \tag{11.8}$$

n is the fine capacitor index, ranging from 0 to $N-1 (N=6)$ as shown in Fig. 11.16. The resolution of C_{eff} is the function of the cycles K over the measurement. This simplified analysis assumes that the out of band noise of the sigma–delta is removed by the FBAR filtering and that the switched capacitors are ideal. This scheme allows very fine capacitance change, without applying small capacitors that have poor process-variation robustness. This fine-tuning method using the sigma–delta modulator has been proposed for *LC*-based and crystal-based digitally controlled oscillators (DCOs) [4, 16], and we apply the technique to the DCFO.

11.5 Measurement

11.5.1 The Prototype Oscillator

The test chip is fabricated by using 90-nm Si CMOS process. Figure 11.17 shows a micrograph and schematic of the prototype DCFO. The chip is mounted on the board and is connected to the FBAR through bonding wires. The area of the DCO core is 280 μm $\times$ 530 μm including a differential inductor for the spurious

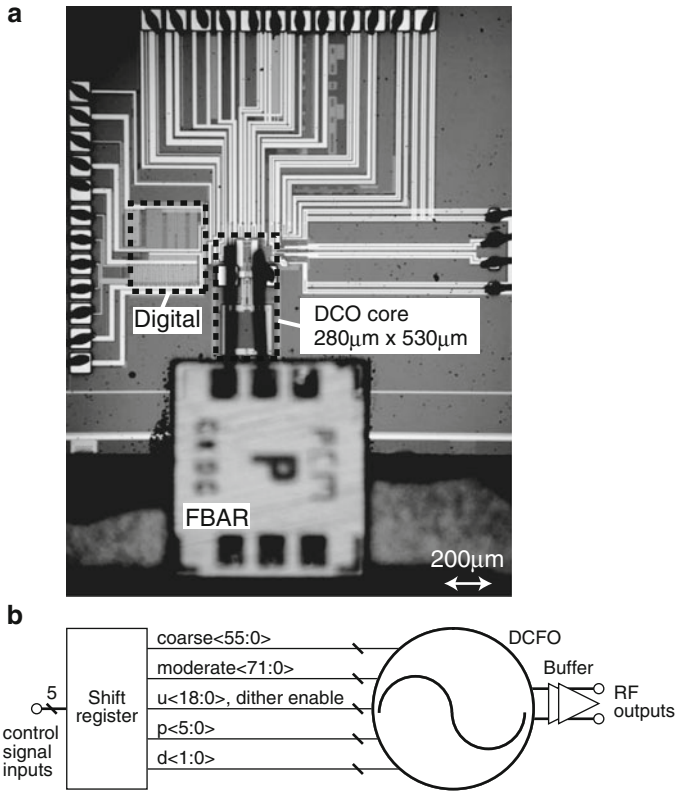

Fig. 11.17 Micrograph and schematic of the prototype in 90 nm CMOS. ©2008 IEEE. Reprinted, with permission, from [13]

suppression. One of the important points for DCFO layout is to suppress series parasitic resistance between the FBAR and cross-coupled transistors M1 and M2: Just a few ohms of series resistance seriously degrades Q of the oscillator tank, i.e., phase noise.

A shift register is implemented into the test chip because of limitation of the IO pad number. The sharp clock transition is preferred to prevent phase noise degradation at buffer circuits, so CMOS-inverter-type buffers are commonly applied for generating the rail-to-rail square wave [4]. However, a class-A amplifier is used in the prototype for estimating spurious signals and quantization noise.

11.5.2 Measurement Results

The measured output spectrum is shown in Fig. 11.18a. A signal analyzer (Agilent, E5052A) is used for measurement. All coarse and moderate capacitors are on, and

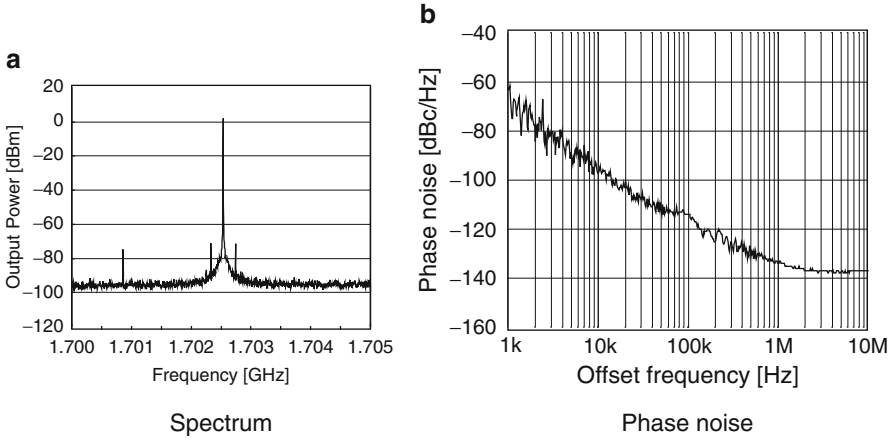

Fig. 11.18 Measured spectrum and phase noise. Numbers of coarse and moderate capacitors are 72 and 56, respectively. The control bit of the sigma–delta modulator u < n > is set to 10,000 in hexadecimal. ©2008 IEEE. Reprinted, with permission, from [13]

fine tuning is also active. Measured output power from the buffer is about 2 dBm. Spurious signals are observed and come from the shift register and equipments; however, they are outside the band of interest. Figure 11.18b is the measured close-in phase noise characteristic. The phase noise at 10 kHz offset is -97 dBc/Hz with the instrument noise floor limiting the noise level over 1 MHz, and the integrated jitter from 1 kHz to 10 MHz is 300 fs.

Figure 11.19 shows the measured frequency tuning characteristics with changing of capacitances. The number of coarse and moderate capacitors is changed. In the proposed oscillator, the oscillation frequency can be controlled over a tuning range of 6.5 MHz from 1,702.7 to 1,709.1 MHz.

Single-sideband modulation is used for measuring resolution of fine bits controlled by the sigma–delta capacitive DAC. A measurement scheme is shown in Fig. 11.20. Output signals of the DCFO are once mixed down because of maximum measurable frequency of a frequency counter. We tested several combinations of control bits for the sigma–delta modulator and several types of periods t_p. A measured resolution per a fine bit was 0.03 ppb.

Performance of the oscillator is summarized in Table 11.4. Power consumption of the DCO core is 1.5 mW, and the sigma–delta modulator consumes 2.3 mW. A figure of merit (FoM) is -200 dBc/Hz at 10 kHz offset. The commonly used FoM is determined by

$$FoM = L\{\Delta f\} - 20\log\left(\frac{f_o}{\Delta f}\right) + 10\log\left(\frac{P_{DC}}{1\text{mW}}\right), \tag{11.9}$$

where $L\{\Delta f\}$ stands for phase noise at a given offset Δf and f_o is the center frequency of the oscillator. P_{DC} is power consumption.

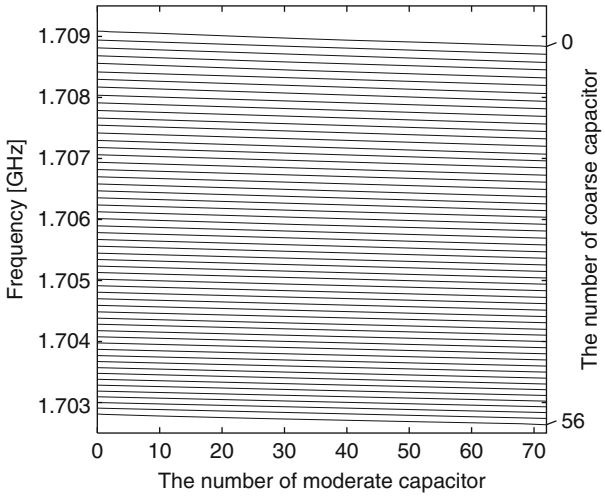

Fig. 11.19 Output frequency as a function of the coarse and moderate tuning codes illustrating the linearity and overlap between banks. ©2008 IEEE. Reprinted, with permission, from [13]

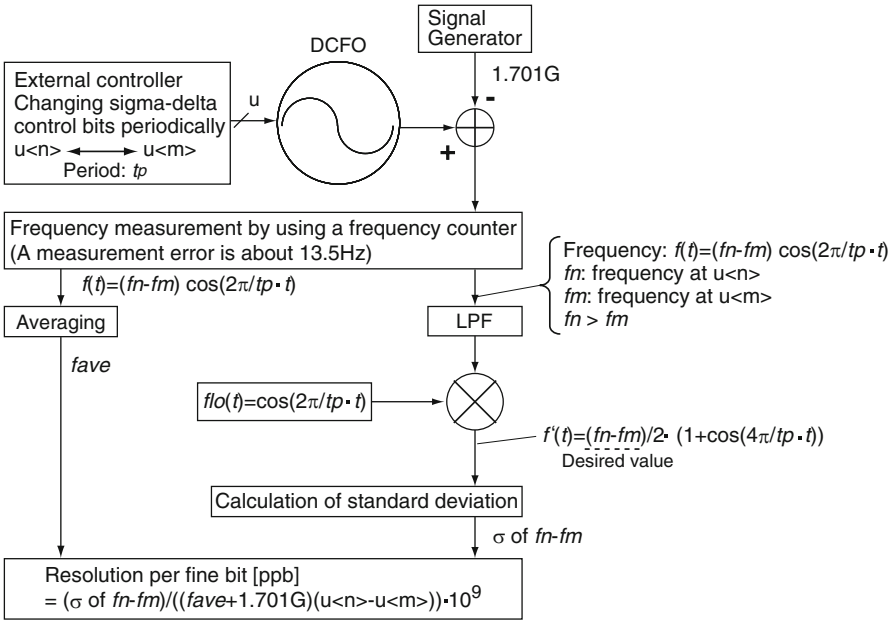

Fig. 11.20 Algorithm for fine capacitor bank resolution measurement. Single-sideband modulation is used. For example, when u < n > and u < m > are set to 10,000 and 08,000 in hexadecimal, u < n > −u < m > becomes 2^{14}. ©2008 IEEE. Reprinted, with permission, from [13]

Table 11.4 Performance summary

Process technology	90-nm Si CMOS
Center frequency	1705.9 MHz
Supply voltage	0.9 V for the DCO core 1.1V for digital circuits
Power consumption	1.5 mW for the DCO core 2.3mW for the sigma–delta modulator
Tuning range	6.4MHz
Phase noise	−97 dBc/Hz at 10 kHz offset
FoM	−200 dBc/Hz at 10 kHz offset
Tuning resolution	0.03 ppb

Table 11.5 Comparison with the present work and previously reported FBAR oscillators

Reference	Process	Power [mW]	Center frequency [GHz]	Phase noise @ 10 kHz offset [dBc/Hz]	FoM [dBc/Hz]	Tuning range [MHz]
This work	90 nm Si CMOS	1.5[a]	1.7	−97	−200	6.4
[1]	0.18μm Si CMOS	0.3	1.9	−100	−211	0
[5]	0.13μm Si CMOS	6.0	2.2	−85[b]	−184	220
[7]	0.13μm Si CMOS	0.6	2.1	−95[b]	−204	1.5
[8]	0.13μm Si CMOS	0.089	1.9	−98	−214	0
[9]	0.35μm SiGe BiCMOS	4.6	5.5	−93	−201	0
[6]	0.25μm SiGe BiCMOS	24.3	2.1	−94[b]	−187	37
[2]	Si bipolar	115.5	2.0	−112	−197	2.5

[a] Only the DCO core

[b] These values are read off from graphs in their papers

Measured phase noise is larger than the target phase noise of −110 dBc/Hz at 10 kHz offset although −111 dBc/Hz at 10 kHz offset has been achieved in our post-layout simulation. Degradation of phase noise was caused by larger-than-expected series parasitic resistance among pads and the FBAR. Difficulties of wire bonding to copper pads on low-k films resulted in increase of contact resistance at the surface between the pad and the bonding wire. However, as shown in Table 11.5, phase noise of our oscillator at 10 kHz offset is small as compared to other frequency-tunable oscillators in CMOS processes. FoM of our DCFO is comparable with other FBAR oscillators. Target frequency tuning range is about 6.8 MHz, and our DCFO could almost cover frequency fluctuation due to manufacturing and temperature variation. Our oscillator achieved a tuning resolution 0.03 ppb, far better than the target resolution of 100 ppb.

11.6 Conclusions

MEMS oscillators have been coming on the market for substituting area- and cost-consuming crystal oscillators. We introduced a DCFO and addressed challenges: frequency changes due to manufacturing and temperature variations and spurious resonance due to interconnects between the oscillator and the FBAR. The oscillation frequency of the FBAR oscillator can be calibrated by using beacon signals from the base station in some RF communication standards. A very high frequency resolution of 0.03 ppb can be achieved by the use of a sigma–delta capacitive DAC. Thus, the DCFO can achieve very high frequency accuracy required by some of the RF communication standards. The FBAR will be placed outside a package of a fully integrated RF system chip because of cost. In that case, spurious resonance induced by inductance of an interconnect path between the chip and the FBAR becomes an issue. However, we can cancel out the spurious by exploiting a negative impedance circuit.

One of the next challenges for FBAR oscillators would be to suppress increase of phase noise due to parasitic resistance between the FBAR and the chip, especially at contact planes of bonding wires and pads on the CMOS chip. For overcoming the challenge, we may also have to investigate not only circuit techniques but also suitable packages for both the DCFO and the RF system considering area, cost, reliability, and so on.

References

1. B.P. Otis, J.M. Rabaey, IEEE J. Solid State Circ. **38**(6), 1271 (2003)
2. A.P.S. Khanna, E. Gane, T. Chong, in *IEEE MTT-S Digest*, 2003, pp. 717–720
3. Ceramic Resonators (CERALOCK) Nov. 17, 2009. http://www.murata.com/products/catalog/pdf/p16e.pdf
4. J. Lin, IEEE J. Solid State Circ. **40**(12), 2726 (2005)
5. P. Vincent, J.B. David, I. Burciu, J. Prouvée, C. Billard, C. Fuchs, G. Parat, E. Defoucaud, A. Reinhardt, in *Digest Paper of IEEE International Solid-State Circuits Conference*, 2008, pp. 478–479
6. K.B. Östman, S.T. Sipilä, I.S. Uzunov, N.T. Tchamov, IEEE J. Solid State Circ. **41**(10), 2248 (2006)
7. S. Rai, B. Otis, in *Digest Paper of IEEE International Solid-State Circuits Conference*, 2007, pp. 576–577
8. Y.H. Chee, A.M. Niknejad, J. Rabaey, in *Proceedings of IEEE Radio Frequency Integrated Circuits Symposium Digest Papers*, 2005, pp. 123–126
9. M. Aissi, E. Tournier, M.A. Dubois, G. Parat, R. Plana, in *Digest Paper of IEEE International Solid-State Circuits Conference*, 2006, pp. 1228–1235
10. J.D.L. III, P.D. Bradley, R.C.R.S. Wartenberg, in *Proceedings of IEEE Ultrasonics Symposium*, 2000, pp. 863–868
11. D. Ruffieux, in *Proceedings of European Solid-State Circuits Conference*, 2002, pp. 85–88
12. E. Hegazi, H. Sjöland, A.A. Abidi, IEEE J. Solid State Circ. **36**(12), 1921 (2001)
13. H. Ito, H. Lakdawala, A. Ravi, S. Pellerano, R. Ruby, K. Soumyanath, K. Masu, in *ESSCIRC 2008. 34th European Solid-State Circuits Conference, 2008*, 2008, pp. 98–101. DOI 10.1109/ESSCIRC.2008.4681801

14. B. Razavi, *Design of Analog CMOS Integrated Circuits* (McGraw-Hill Higher Education, New York, 2001)
15. 2007 International Technology Roadmap for Semiconductors. Table RFAMS1a. http://www.itrs.net/Links/2007ITRS/2007_Chapters/2007_Wireless.pdf
16. R.B. Staszewski, D. Leipold, K. Muhammad, P.T. Balsara, IEEE Trans. Circ. Syst. II **50**(11), 815 (2003)
17. R. Schreier, G.C. Temes, *Understanding Delta-Sigma Data Converters* (IEEE, New Jersey, 2005)

Chapter 12
A Robust Wireless Sensor Node for In-Tire-Pressure Monitoring

Markus Dielacher, Martin Flatscher, Thomas Herndl, Thomas Lentsch, Rainer Matischek, Josef Prainsack, and Werner Weber

Abstract In this chapter, we describe the architecture of an in-tire-pressure sensing system for automotive applications. Challenges are the supply of sufficient power, the proper functioning of the sensors (pressure, acceleration, and temperature), and a transceiver architecture with extremely low power consumption. Power consumption is reduced by use of a high-frequency resonator based on the bulk acoustic wave (BAW) technology. Such devices are used both in the transmit and in the receive paths. Further power-saving measures are taken introducing a specially adapted on/off cycling operation. The power supplied to the system is provided either by two small batteries or an energy scavenger, both capacitor buffered. The electronic system is integrated into 3D stacks using either through-silicon vias or the ultrathin chip stacking method to provide short interconnects for further reduced power consumption. The overall system is mounted in a compact molded interconnect device to save space and weight.

12.1 Power Aware System Architecture

12.1.1 Application Scenario

Today's cars are equipped with an increasing number of sensor devices which are in most cases connected via cables. However, there is an increasing demand

M. Dielacher (✉) • M. Flatscher • T. Herndl • T. Lentsch • R. Matischek • J. Prainsack
Infineon Technologies, Graz, Austria
e-mail: Markus.Dielacher@infineon.com; martin.flatscher@infineon.com; thomas.herndl@infineon.com; thomas.lentsch@infineon.com; rainer.matischek@infineon.com; josef.prainsack@infineon.com

W. Weber
Infineon Technologies, Munich, Germany

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3_12,

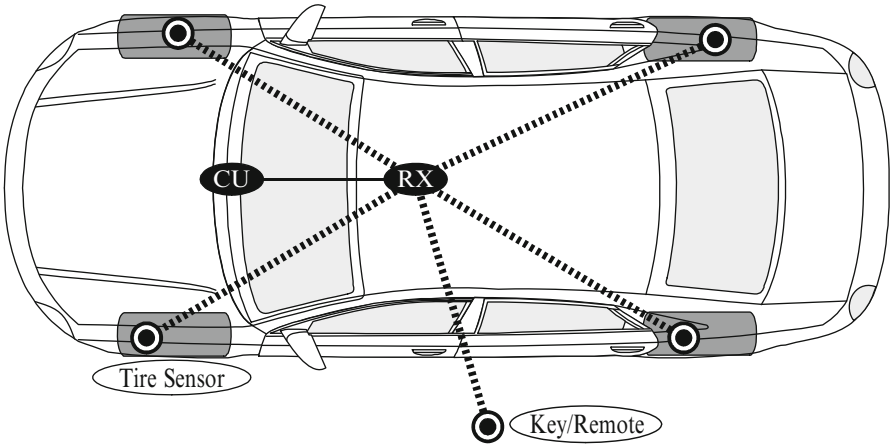

Fig. 12.1 State-of-the-art wireless sensor network in a car

for wireless applications. Currently, they are used for scenarios where wire-based communication is simply not possible, such as tire pressure sensors mounted in the wheel or remote keyless entry (RKE) transmitters. In the future, there will be a need to extend that wireless sensor network with sensors that are still wired today in order to reduce the amount of cables and gain higher flexibility. This chapter focuses on the first step in that direction, the optimization of wireless sensor nodes for the existing wireless applications.

A typical application scenario of a state-of-the-art wireless sensor network in a car is sketched in Fig. 12.1. The tire pressure sensors as well as the electronic car key (or an advanced remote control) transmit an individual identifier code and additional data payload to a common receiver (RX) in the central unit (CU). Optionally the data transmission is repeated in order to increase the probability that the message is not disturbed by interference or other radio noise.

12.1.2 System Power Optimization

In order to reduce the power consumption of the entire wireless network, the communication protocols have to be optimized for low-power operation. This is quite evident for the battery-powered tire pressure sensors and the RKE devices because they have to operate self-sufficiently for several years. However, it is also important that the power consumption of the wireless receiver system is minimized, despite it is powered by the high-capacity vehicle electrical system. The reason is that the receiver has to be active for the whole time the car is parked and wait for a key to transmit. In this case, it has to be ensured that the receiver system does not drain the car battery even when parking is for some weeks or months.

An efficient way to reduce the overall power consumption of the wireless communication system without decreasing the reception quality is to use duty cycling of the radio units. For applications with cyclically transmitting sensor nodes, there already exist optimized medium access control (MAC) protocols [1,2]. In these protocols, the sensor nodes exchange their schedules during runtime and only wake up at predefined points in time in order to reduce the power consumption for idle listening.

However, such scheduled protocols cannot be used for tire pressure monitoring systems (TPMS) or RKE. In these applications, the sensor nodes do not only transmit cyclically but also immediately at random events. Such an event could be a loss of pressure, a sudden change of acceleration, or simply the unlocking of the doors by the driver. Therefore, another MAC variant is used. The basic principle is similar to the common "Low Power Listening" protocol [3]: The radio reception is activated for a short time followed by a longer period where the radio is in a power down mode and reception is disabled respectively. During the reception period, the advanced systems are first listening for a carrier and only in case of a detected signal they keep listening for a specific header or preamble. The downside of this duty-cycled receiver protocol is that if the transmission is acyclic, the transmitting sensor nodes have to send a longer preamble (higher overhead) matched with the duty cycle of the receiver. Consequently, the following trade-off has to be considered: On the one hand, the overall power consumption of the receiver unit is reduced, but on the other hand, the longer preamble comes along with increased power consumption for each transmission. In case of the tire pressure sensor nodes, the typical duty cycle of the data transmission is extremely low (a few ms transmission every minute). Therefore, the transmission of an extended preamble is an acceptable trade-off.

However, the energy saving by duty cycling with conventional low-power radios is rather limited, since they are typically based on crystal oscillators. The main limiting aspect is the start-up time of the crystal and the PLL which is typically in the range between 0.5 [4] and 1 ms [5] or even longer. In the case of the data rates of typical low-power wireless sensor networks (20–50 kbps), the actual time for carrier sensing (or receiving a short preamble) is about 0.3–0.6 ms; thus, the crystal start-up time can become dominating.

In order to overcome these limitations, the presented sensor node utilizes a novel MEMS-based radio architecture which allows for a significantly faster start-up of the oscillator. The radio uses a free running bulk acoustic wave (BAW) [6] oscillator that can be fine-tuned digitally and thus avoids the need of a crystal-based PLL architecture. It achieves a start-up time of about 2 µs which is up to 1,000 times faster than crystal-based radios. This feature is especially beneficial for duty-cycled wireless sensor networks applying idle listening: A shorter start-up time can be used for a higher repetition rate with a constant duty cycle at the receiver. In this case, only a shorter preamble is required, which especially reduces the overhead and power consumption of the transmitting nodes. Another advantage

of this configuration will be important for future wireless sensor applications: The shorter start-up time and shorter cycle time of the receiver allows for a significantly decreased message delay independent of the used MAC protocol, so it can be used for more time-critical wireless applications.

12.2 Application of In-Tire-Pressure Monitoring

State-of-the-art tire pressure monitoring systems are wireless sensor nodes mounted on the rim (see Fig. 12.2).

Attaching the node on the inner liner of a tire allows sensing of important additional technical parameters, such as road condition, tire wearout, temperature, tire friction, side slip, wheel speed, and vehicle load. These may be used for improved tracking and engine control, feedback to the power train, and car-to-car communication purposes. Thus, a significant step in car stability control appears feasible. However, this is not straightforward because of the constraints on the weight and size of the node. The maximum weight of the sensor is limited to 5 g including package, power supply, and antenna. The node size is limited to about $1\,\mathrm{cm}^3$ to avoid high force gradients due to device deformation. Robustness is required against extreme levels of acceleration of up to 3,000 g, and beyond ($g = 9.81\,\mathrm{m/s}^2$) and finally, a long lifetime of the module must be achieved.

The presented, self-sufficient, tire-mounted, wireless sensor node (Fig. 12.3) has been:

- Designed with a BAW-based low-power FSK transceiver.
- Prepared for an energy scavenger-based low-volume and low-weight power supply.
- Arranged in a 3D vertical chip stack for best compactness, lowest volume, and highest robustness for pressure, inertia, and temperature sensing.

All in all, the sensor node includes four different dies: a sensor, a microcontroller, a radio transceiver, and a BAW die. The sensor die is a MEMS device containing sensors for pressure and acceleration. The sensor interface together with an ADC for converting the sensor data is integrated on the microcontroller. Besides that,

Fig. 12.2 State-of-the-art TPMS

Fig. 12.3 In-tire sensor node

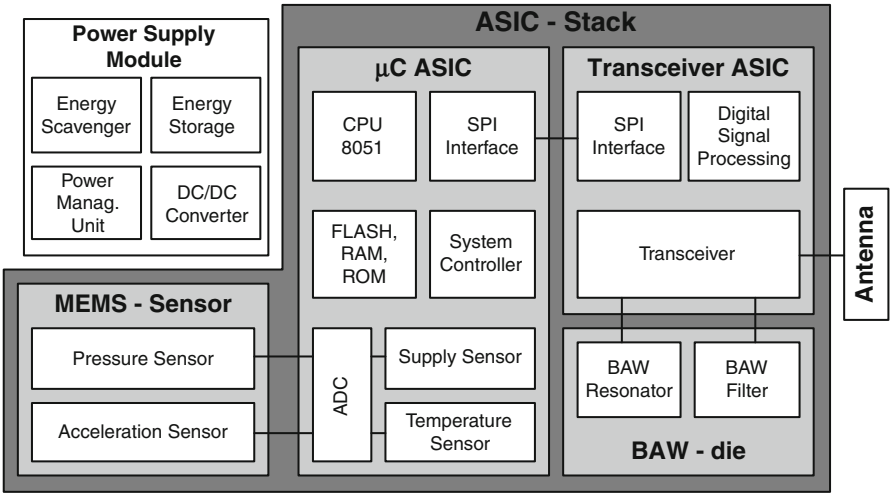

Fig. 12.4 Basic block diagram of the wireless sensor node. ©2010 IEEE. Reprinted, with permission, from [7]

the microcontroller contains RAM, ROM, and also a FLASH memory. It provides additional sensors for temperature and the battery voltage. The communication with the transceiver ASIC is established via a serial peripheral interface (SPI). The BAW die contains separate resonators, two of which are connected to oscillators on the transceiver ASIC and further two are used as filters. The basic block diagram of the wireless sensor node is illustrated in Fig. 12.4.

In order to supply the sensor node with energy, a power supply module is integrated. In a first version, this power supply module contains small batteries together with capacitors which are required because the batteries are not able to

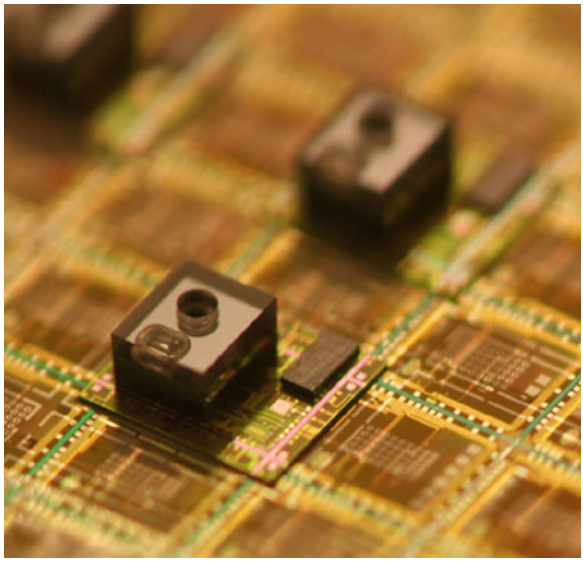

Fig. 12.5 ASIC stack (Source: SINTEF). ©2010 IEEE. Reprinted, with permission, from [7]

provide the required peak current during transmission. The batteries slowly charge the capacitors which are then used to provide the desired peak current.

In a second version of the power supply module, an energy scavenger converts vibrations into electrical energy, which is stored in large capacitors again. For the energy scavenger, it is very beneficial to attach the sensor node to the inner liner of the tire instead of the rim because there is much more vibration energy available in that position.

The microcontroller, the transceiver, the sensor, and the BAW dies are arranged in two different versions of 3D chip stacks using either through-silicon vias (TSV) [8] or the ultrathin chip stacking (UTCS) method [9]. This integration reduces the required connections to external components to a minimum, and only very few additional wire bonds are necessary. The whole ASIC stack is bonded onto a PCB which is then mounted into a molded interconnect device (MID). Beside the ASIC stack, the PCB only contains a few external capacitors for voltage regulators and to fine-tune the resonant frequency of the antenna. Figure 12.5 shows the chip stack using TSVs. The microcontroller is the bottom layer of the stack. It is carrying the transceiver which has been mounted by making use of μBumps. On top of the transceiver, one can see the sensor and the BAW die, both flip-chipped. In Fig. 12.5, the dicing of the microcontroller wafer remains to be done.

The MID package consists of a modified polymer, which is appropriate for injection molding. The modified polymer allows for laser activation of circuit tracks on the surface of the device. After laser activation of the circuit tracks, the MID is put into a chemical bath in a separate process step; then only the activated areas

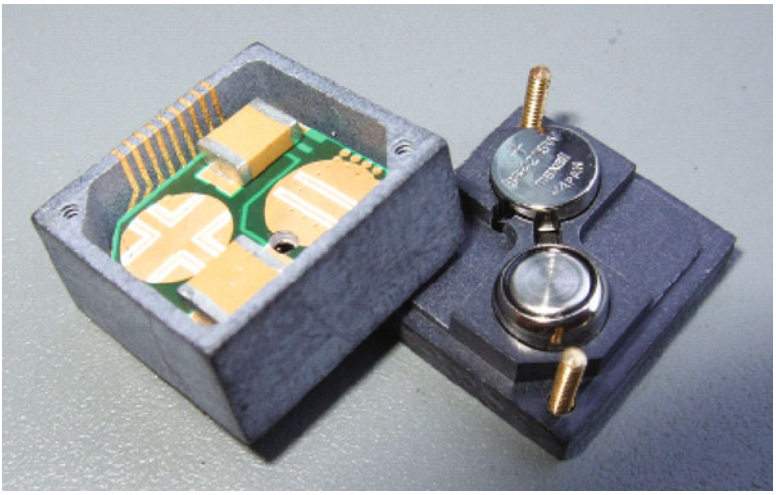

Fig. 12.6 Integrated sensor node

get metalized. A loop antenna is also fabricated in that way on the inside walls of the MID. The whole sensor node has a volume of about 1 cm^3 and can be seen in Fig. 12.6.

12.3 Power Aware Radio Architecture

In Fig. 12.7, the schematic of the BAW-based transceiver is shown. It contains two fully integrated BAW oscillators, one is used for generating the modulated carrier in transmit mode and the other one provides a local oscillator (LO) signal in receive mode. The two oscillators have a frequency offset for realizing a heterodyne receiver with a low intermediate frequency (IF) of 10.7 MHz. The transceiver has been evaluated with two different BAW devices. The resonant frequency of the BAWs is around 2.1 GHz in the first version and 2.45 GHz in the second version. The oscillators are tunable over a certain frequency range to compensate for temperature and process variation, as well as frequency modulation. The tuning can be done via a digitally controlled capacitor bank in parallel to the resonators or with a DC voltage. In most cases, the tuning range of the BAW oscillators will not be sufficient to provide sufficient channel selection. Due to the low costs and due to the fact that the BAW devices can easily be integrated, it might be a reasonable solution to use a bank of BAW resonators with slightly different resonant frequencies as an alternative.

The oscillators have a very short turn-on time, and this fact in combination with the small data packet of the sensor node results in a very narrow active time slot, and therefore, the system is very energy efficient. The short turn-on time is feasible

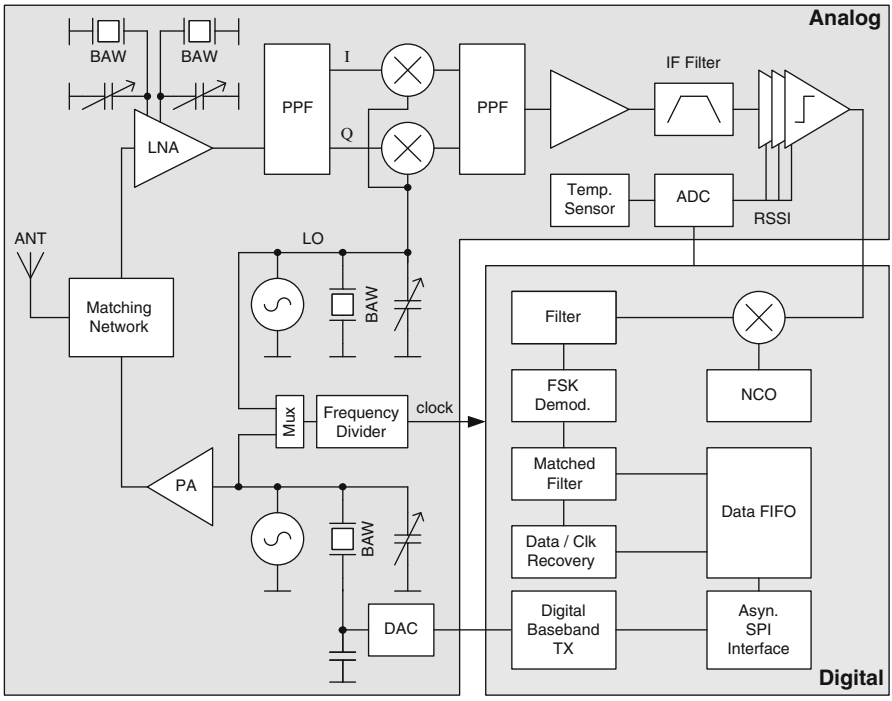

Fig. 12.7 BAW-based transceiver

because the reference frequency is already at RF and the turn-on time is proportional to the quality factor multiplied by the period. In transmit mode, the modulated carrier is fed to a power amplifier after preamplifying. The power amplifier is able to apply a power of about 1 dBm into a 50 Ω load.

The chosen receiver structure is an image-reject architecture. An on-chip matching network, containing also the RX/TX switch feeds the incoming radio signal to a differential LNA, which also includes two BAW resonators for filtering purposes. Usually, the I and Q phases, required for the image-reject mixers are generated by the local oscillator. The applied BAW oscillator, however, does not provide quadrature phases because a quadrature oscillator would have a higher current consumption. That is why the I and Q phases are generated in the signal path by the polyphase filter (PPF) which follows the LNA.

After downconversion of the signal by the mixers, another PPF is responsible for the image rejection. After filtering the desired channel out of the downconverted band, the signal is fed into a limiter with a gain of more than 80 dB. The limiting amplifier eliminates the need for an automatic gain control and delivers a binary signal. Additionally, the limiting amplifier provides an RSSI (received signal strength indicator) signal, which is converted to a digital codeword by a 10 bit ADC. The RSSI signal can be used as wake-up criterion as well as for fine-tuning of the

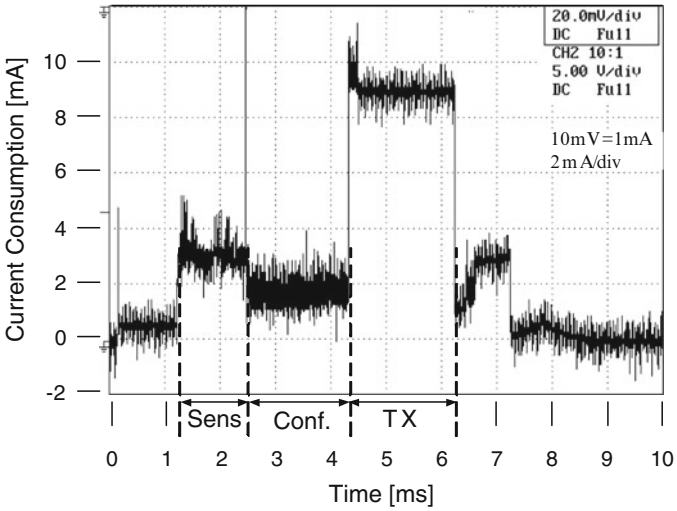

Fig. 12.8 Current profile for one event, measured with a 10 Ω shunt resistor. ©2010 IEEE. Reprinted, with permission, from [7]

LO frequency, the BAW filters, and the matching network. In the digital domain, the binary output of the limiter is directly mixed into complex baseband by using a numerically controlled oscillator (NCO), which allows tuning of the frequency in very fine frequency steps. The digital clock can be derived from any of the two oscillators. After additional digital filtering and demodulation, the signal is passed to the matched filter. By means of the clock and data recovery unit, the output of the matched filter is sampled, and the received payload is stored in a FIFO memory, where the microcontroller can access the data via the SPI. In order to access the FIFO and to allow configuration of the transceiver while the oscillators are turned off, the SPI is implemented asynchronously.

Figure 12.8 shows a typical timing diagram of a BAW-based TPMS sensor node. The duration of one complete reporting event is about 7 ms, and the total required charge is about 30 µC. Even when considering one reporting event every 10 s (only required during alert, when the tire pressure is critical), this contributes to only 3 µA to the overall current. To obtain the overall current consumption, the standby current has to be taken into account. The main part of the standby current is caused by the RC oscillator in the system's wake-up unit (1 µA) and leakage of the monolithic ceramic buffer capacitors. In the presented prototype, the time which is required for temperature compensation and configuration of the transceiver before RF transmission is determined by the slow communication between the microcontroller and the transceiver due to the slow clock of the microcontroller. The transmitter is able to transmit at very high data rates because the BAW oscillator can directly be modulated with a DC voltage.

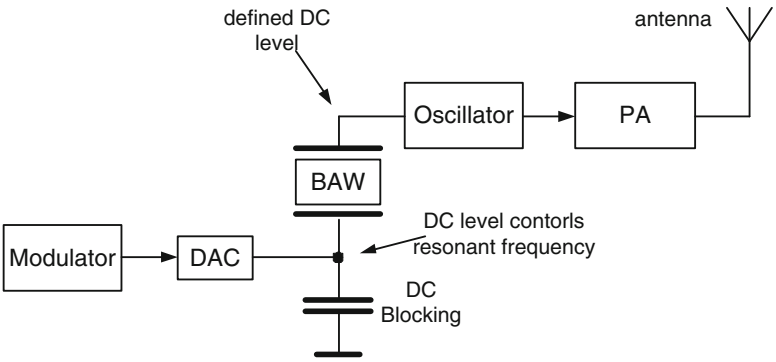

Fig. 12.9 Schematic of the transmitter including BAW oscillator

12.3.1 BAW-Based Transmitter

Figure 12.9 shows a more detailed schematic of the transmit section of the transceiver, including the BAW oscillator together with the circuit for frequency modulation via variation of the DC biasing. In addition to the tuning with the digitally controlled capacitor bank, the BAW can be tuned with a variable DC voltage, which causes a change in the stiffness of the piezoelectric BAW material. The biasing voltage, generated by a 5 bit DAC, is applied at the bottom electrode (AC ground) of the resonator and causes a linear frequency shift with a slope of 40 kHz/V. The applied resonators are mirror-type BAWs, also called solidly mounted resonators (SMR) [10]. Typically, the parasitic capacitances of the electrodes of such BAWs are not equal, since the bottom electrodes have a higher capacitance against substrate. The advantage of connecting one electrode to AC ground is that the output impedance of the DAC does not have an influence on the digitally controlled capacitor bank. Besides that, the required DAC can easily be implemented into the chip, because the required resolution is only in the range of a few bits and consumes only little additional chip area. This architecture is capable of providing very high data rates at a very low complexity.

One critical design issue is the fact that the impedance of the BAW resonator is relatively low. Particularly when loading the resonator with a huge capacitance (e.g., tuning capacitors, ESD protection circuit, and parasitics), the impedance at parallel resonance can drop below 500 Ω. To guarantee start-up of oscillation, the transconductance of the oscillator core must be high enough. Because of the high Q factor of the resonator (1,000–2,000), the phase shift of the oscillator core has to be very low while keeping the gain high. In Fig. 12.10, the basic principle of the BAW oscillator is shown [11]. Assuming the sources of transistors M1 and M2 in Fig. 12.10 are connected to ground, the signal on node A amplified by M1 results in a 180° shifted signal on node B. To avoid any additional phase shift, node B should not be capacitively loaded, and hence, a simple resistor is used as load element.

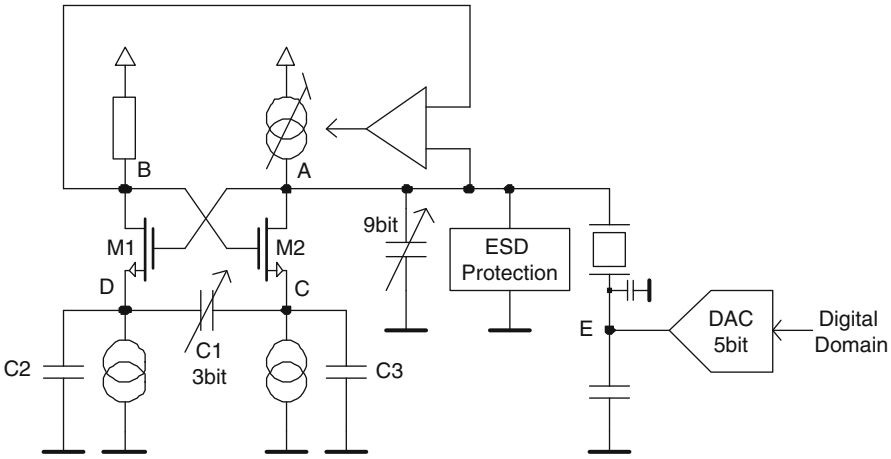

Fig. 12.10 Basic principle of the BAW oscillator. ©2010 IEEE. Reprinted, with permission, from [7]

The capacitive load of node A is not so critical because it only causes a shift of the BAW resonance frequency and degrades the impedance as already discussed. In order not to damp the high Q resonant circuit, the resonator should be connected to a high-impedance node. Therefore, a current source instead of a resistor is used as load element of transistor M2. With the sources of M1 and M2 connected to ground, the circuit obviously would not oscillate but latch and nodes A and B would saturate. Therefore, the combination of C1, C2, and C3 together with two further current sources has been introduced for DC decoupling. With C1, the phase shift can be trimmed, which can also be used for tuning the frequency in a very small range with 3 bit resolution. The high gain of the oscillator core could cause stability problems in the control loop of the current source, operating as load of transistor M2. In order to maintain stability more easily, a replica circuit (not shown in the schematic) without the cross coupling of transistors M1 and M2 is applied.

Another critical design issue is the temperature drift of the BAW resonant frequency of about -18 ppm/°C. To overcome this temperature drift, the temperature is measured and compensated via the binary weighted digitally controlled capacitor bank with a resolution of 9 bit in parallel to the resonator in the range of -40 °C to +125 °C by applying a simple temperature compensation algorithm.

The impacts of potential heat sources, the environment, and the active areas on the dies have been investigated thoroughly. Figure 12.11 depicts temperature variations of a typical transmission cycle [12]. In order to keep the required peak current small, temperature measurement and RF transmission are never performed at the same time—but in different phases of the transmit cycle. Furthermore, the temperature sensor is located at the center of the transceiver die, and the BAW die is flip-chipped on the transceiver using gold stud bumps. Thus, there might appear a temperature difference at the different locations as well as a temperature difference

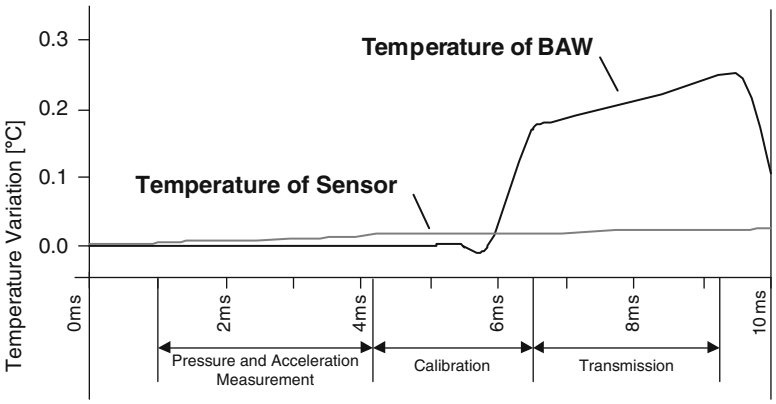

Fig. 12.11 Temperature variations during one transmission cycle on BAR die and temperature sensor

at different points in time. Every phase of the cycle causes heat sources to appear on different locations distributed all over the 3D stack at different points in time. Therefore, a certain measurement error must be taken into account.

Assuming only 1/°C temperature difference between the temperature sensor in the measurement phase and the BAW die in the RF transmission phase, the resulting shift of the carrier frequency in the transmitter is 38 kHz due to the drift of 18 ppm/°C. The same might happen in the LO at the receiving side leading to a 76-kHz frequency offset at the intermediate frequency (IF), where the channel filtering is implemented. For applications such as TPMS, a bandpass filter with a narrow bandwidth of about 300 kHz is advantageous, because the bandwidth is sufficient for the low data rate while interferers and noise are rejected. The deviation of the center frequency reduces the bandwidth twice, because it can be shifted toward the lower as well as to the upper corner frequency of the filter, resulting in an effective bandwidth of:

$$BW_{eff} = BW - 2 \cdot \Delta f_c = 300\,\text{kHz} - 2 \cdot 76\,\text{kHz} = 148\,\text{kHz}. \tag{12.1}$$

Hence, the frequency offset caused by only 1/°C temperature difference between sensor and BAW die reduces the available bandwidth for communication to only one half of the filter bandwidth. On the one hand, the data rate needs to be reduced in order to enable the received signal to pass the channel filter undisturbed, and on the other hand, the signal-to-noise ratio (SNR) is decreased due to the large unused part of the filter bandwidth.

Therefore, the understanding of temperature distributions and displacements is crucial for the performance of the sensor node. While the appearance of external heat sources is more or less random and can only be taken into account in terms of worst-case scenarios, the temperature drift due to self-induced power dissipation within the 3D stack is well predictable and can be compensated. For that reason,

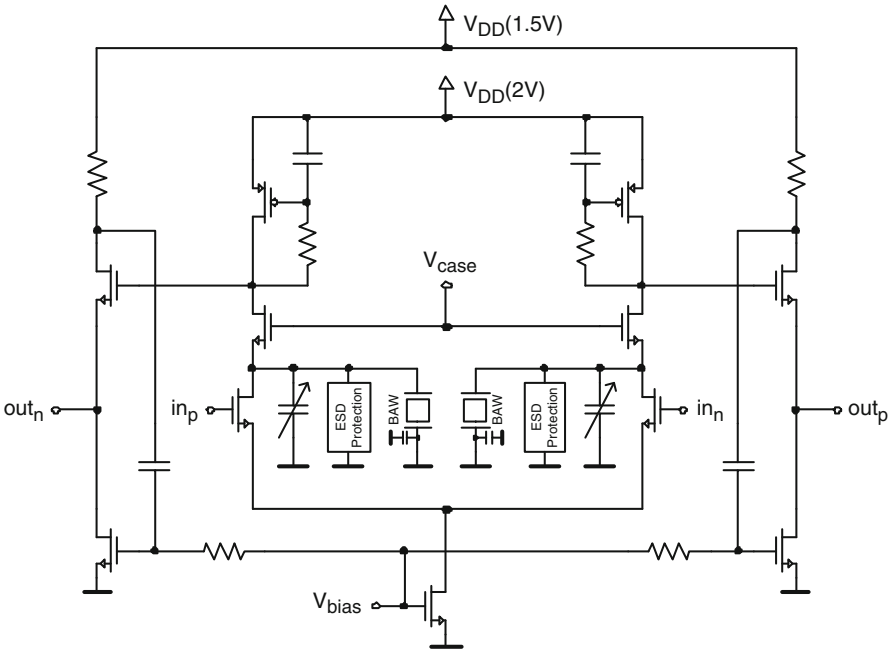

Fig. 12.12 Differential LNA with integrated BAW resonators

manifold simulations, have been performed [12]. For the simulations the location, power dissipation, and timing of all heat sources distributed on the three different dies (BAW, transceiver, and microcontroller) have been defined exactly.

According to the results of those simulations, the maximum temperature difference between the sensor during calibration and the BAW during transmission can be estimated to be +0.25/°C when constant temperature around the sensor node is assumed. Neglecting this measurement error during calibration would cause a systematic frequency shift of 4.6 ppm.

12.3.2 BAW-Based LNA

The schematic of the filtering LNA is shown in Fig. 12.12. It consists of a differential common source input stage with cascode transistors, and an active inductor stage is used as load. The source follower output stage includes a feedback loop to decrease the output impedance.

As mentioned above, the impedance of the BAW resonators is quite low, even at their parallel resonant frequency, where the impedance is at a maximum. That is why it is not possible to insert the BAWs at high-impedance nodes in the LNA,

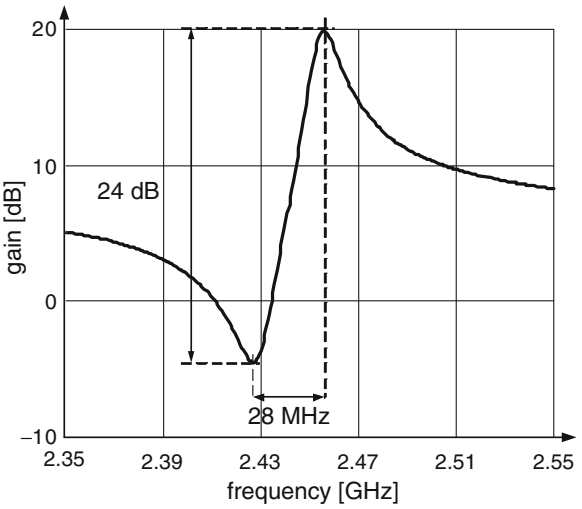

Fig. 12.13 Image suppression in the LNA. ©2009 IEEE. Reprinted, with permission, from [13]

for example, in parallel to the load; otherwise; the gain of the circuit would degrade significantly. In the presented circuit, the nodes at the input of the cascode have been chosen for inserting the BAWs because they are at a relatively low impedance level. The reason for using two BAWs instead of only one connected in between the two branches of the differential circuit is the higher capacitance against substrate in one of the two electrodes, which makes the BAW nonsymmetrical.

In addition to the image-reject architecture, the BAW filters in the LNA can be used to suppress the image frequency significantly. Figure 12.13 shows the simulated voltage gain of the LNA. The difference between the series and the parallel resonant frequency of the BAW is only 28 MHz or even less, if the BAW is tuned. The choice of an intermediate frequency of 10.7 MHz is made, because in that way, it is possible to place the image frequency very close to the series resonance while the desired signal is at the parallel resonance with the local oscillator in between and due to the availability of cheap ceramic filters for that frequency. The plot in Fig. 12.13 does not include the gain of the matching network.

As the presented LNA does not include any measures for impedance matching like inductive source degeneration, its input impedance is almost purely capacitive. This fact is exploited by the on-chip matching network which uses the capacitive LNA input together with an on-chip inductor to generate a resonant circuit. The losses of this resonant circuit are then matched to the 50 Ω antenna. In this way, the matching network accounts for another $\sim$10 dB of voltage gain. A detailed description of the matching network can be found in [13]. Figure 12.14 shows the combined gain and noise figure of the LNA and the matching network. When the filter BAWs in the LNA are not tuned at all, then the total gain is 31.5 dB at a noise figure of 4.6 dB. With maximum tuning, the gain drops to 28.3 dB at a noise figure of 5.4 dB. The current consumption of the LNA is 2 mA.

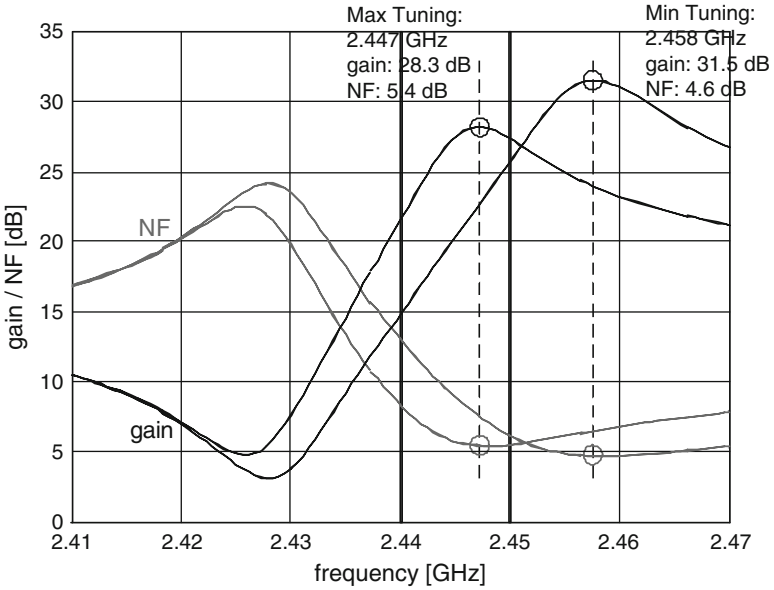

Fig. 12.14 Simulated voltage gain and NF of LNA and matching network. ©2009 IEEE. Reprinted, with permission, from [13]

12.4 Performance Summary

A fully self-sufficient tire pressure monitoring sensor node is presented, which applies BAW resonators for frequency generation instead of using a classical PLL architecture. With this approach, it is possible to reduce the turn-on time of the transceiver by a factor of up to 1,000 compared to conventional crystal-based PLL systems, and therefore, the energy efficiency increases significantly. The whole sensor node has a volume of about 1 cm^3, which is feasible by integrating the different dies into a 3D chip stack and mounting it into a MID. The node includes sensors for pressure, acceleration, temperature, and the battery voltage. The transceiver operates at either 2.1 or 2.45 GHz depending on the assembled BAW devices and consumes 6 mA in transmit mode at 1 dBm output power and 8 mA in receive mode. The sensitivity of the receiver is −90 dBm at a data rate of 50 kBit/s and a bit error rate of 10^{-2}.

References

1. A. El-Hoydi, J.D. Decotignie, in *Proceedings of the International Symposium on Computers and Communications (ISCC)*, 2004, pp. 244–251
2. C.C. Enz, A. El-Hoiydi, J.D. Decotignie, T. Melly, V. Peiris, IEEE Comput. Mag. **37**(8), 62 (2004)

3. J. Polastre, J. Hill, D. Culler, in *SenSys*, 2004
4. TDA5250 D2 ASK/FSK 868MHz Wireless Transceiver. Data sheet, Infineon Technologies AG (2007) http://www.infineon.com
5. TDK5100 ASK/FSK 868/433 MHz Wireless Transmitter. Data sheet, Infineon Technologies AG (2002) http://www.infineon.com
6. F.Z. Bi, B.P. Barber, IEEE Microw. Mag. **9**(5), 65 (2008)
7. M. Flatscher, M. Dielacher, T. Herndl, T. Lentsch, R. Matischek, J. Prainsack, W. Pribyl, H. Theuss, W. Weber, IEEE J. Solid State Circ. **45**(1), 167 (2010). DOI 10.1109/JSSC.2009.2034436
8. P. Ramm, A. Klumpp, R. Merkel, J. Weber, R. Wieland, A. Ostmann, J. Wolf, in *MRS Symposium Proceedings 766*, 2003, pp. 3–14
9. E. Beyne, in *Handbook of 3D Integration: Technology and Applications of 3D Integrated Circuits*, vol. 2 (Wiley, Weinheim, 2008)
10. R. Aigner, in *IEEE Custom Integrated Circuits Conference*, 2003, pp. 141–146
11. M. Flatscher, M. Dielacher, T. Herndl, T. Lentsch, R. Matischek, J. Prainsack, W. Pribyl, H. Theuss, W. Weber, in *International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2009, pp. 286–287
12. J. Prainsack, M. Dielacher, M. Flatscher, T. Herndl, R. Matischek, J. Stolle, W. Weber, in *Design, Test, Integration & Packaging of MEMS/MOEMS (DTIP)*, 2009, pp. 81–85
13. M. Dielacher, M. Flatscher, W. Pribyl, in *Research in Microelectronics and Electronics PRIME*, 2009, pp. 172–175

Index

C.C. Enz and A. Kaiser (eds.), *MEMS-based Circuits and Systems for Wireless Communication*, Integrated Circuits and Systems, DOI 10.1007/978-1-4419-8798-3,