

Circuits and Systems for Wireless Communications

Edited by

Markus Helfenstein

and

George S. Moschytz

Swiss Federal Institute of Technology, Zurich

KLUWER ACADEMIC PUBLISHERS

NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

EX-1046

Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

eBook ISBN: 0-306-47303-8
Print ISBN: 0-792-37722-2

©2002 Kluwer Academic Publishers
New York, Boston, Dordrecht, London, Moscow

All rights reserved

No part of this eBook may be reproduced or transmitted in any form or by any means, electronic, mechanical, recording, or otherwise, without written consent from the Publisher

Created in the United States of America

Visit Kluwer Online at: <http://www.kluweronline.com>
and Kluwer's eBookstore at: <http://www.ebooks.kluweronline.com>

3 INTEGRATION OF RF SYSTEMS ON A CHIP

Peter J. Mole

Nortel Networks
London Road
Harlow, Essex, CM 17 9NA, U.K.

3.1 RF ISSUES

Before considering the details of implementation of an RF system on a chip, it is worth spending some time discussing the problems that any radio system must overcome if to achieve an acceptable performance. The important aspect to remember is that the radio system never operates in isolation. It is not sufficient to think only of the transmitter, the link and the receiver, though many problems lie in this simple chain alone. It is also important to realise that the transmitter can interfere with other links, and a receiver may be unduly sensitive to unwanted, but entirely legitimate, signals.

3.1.1 *Receiver concerns*

In any system, a band is defined in which the receiver may receive signals. The receiver must be able to reject signals outside the band without loss of performance. These signals may be very large (e.g. TV transmissions) and quite capable of overloading a sensitive receiver. If the receiver is overloaded, amplifier compression will reduce the receiver gain and hence the ability to detect weak signals, or the non-linearities in the receiver that are excited by the overload will allow unwanted signals to intermodulate. This may result in a distortion product falling onto the wanted sig-

23

nal and effectively masking it. It is therefore imperative that these signals are heavily filtered before the input to the receiver.

The receiver must respond to any signal in the system band, however, it must also reject unwanted in-band signals without suffering from overload. Such unwanted signals may be significantly larger than the wanted; consider the situation when you are trying to make a cellular phone link to your service provider's remote base-station but, unfortunately, you are physically adjacent to a second service provider's base-station. The system specification provides guidelines to the levels of signals which can exist. Fig. 3.1 shows the relative levels for signals in the DECT cordless phone system. In this system, where communication channels are allocated to both time and frequency slots, the frequency channels are separated by 1.728 MHz. A weak signal, (defined in the DECT specification as -73 dBm) must be receivable in the presence of a signal 13 dB stronger in the neighbouring channel and 34 dB stronger in the next neighbouring channel. To achieve this rejection of neighbouring interferers, a tunable filter is required. In practice this is achieved by mixing the signal to a fixed intermediate frequency and then filtering. The specifications reflect the fact that it is more difficult to filter the neighbouring channel than more distant ones. Until the signals can be filtered, however, the radio must be designed so that the interferers do not create overload. Unlike out-of-band blocking signals, in-band interferers cannot be filtered prior to the sensitive input of the receiver.

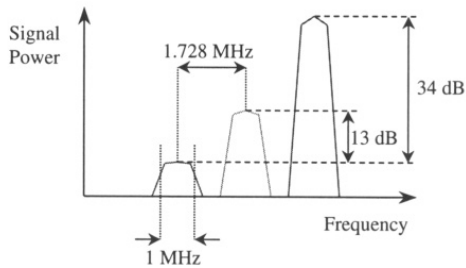


Figure 3.1 Adjacent channel interferers for DECT system.

The frequency translation that is required in a receiver to move the radio frequency input to the required intermediate frequency for filtering is achieved with a mixer. A mixer effectively multiplies the RF input signal with a chosen local oscillator (LO) signal. Thus we can ideally represent the RF input by $\sin(\omega_r t)$ and the local oscillator by $\cos(\omega_L t)$. Thus the output can be written as:

$$\sin \omega_r t \cdot \cos \omega_L t = 0.5(\sin(\omega_r + \omega_L)t + \sin(\omega_r - \omega_L)t).$$

From this we can see immediately that an unwanted output is produced at the sum frequency, but this is easily filtered because of the frequency difference. A more subtle defect is that two input frequencies ($\omega_L \pm \omega_{IF}$) can produce signals at the intermediate frequency (ω_{IF}). Thus there are two frequencies we are sensitive to, the wanted and the

image frequency. It is therefore essential that any signal present at the image frequency is filtered out in front of the mixer.

In practice, to achieve efficient mixing, it is normal that the LO signal is close to a square wave. Thus the LO signal will contain the third harmonic of the LO (about 9 dB lower in signal amplitude) which is also multiplied by the input signal. This immediately means that signals at $(3\omega_L \pm \omega_{IF})$ will also appear at the output as an intermediate-frequency signal. Other harmonics will behave similarly. Again input signals at these frequencies must be removed prior to mixing.

Another defect of the LO signal is that it suffers from phase noise (or jitter). This means the LO is not a pure tone but a spectrum of signals centred on the desired frequency. Thus the mixer will respond to signals close to the wanted signal with a sensitivity falling off with the phase-noise spectrum. This problem is known as reciprocal mixing. Thus, to ensure that the receiver can reject neighbouring channels, it is important to provide an LO signal with low phase noise. These issues are illustrated in Fig. 3.2.

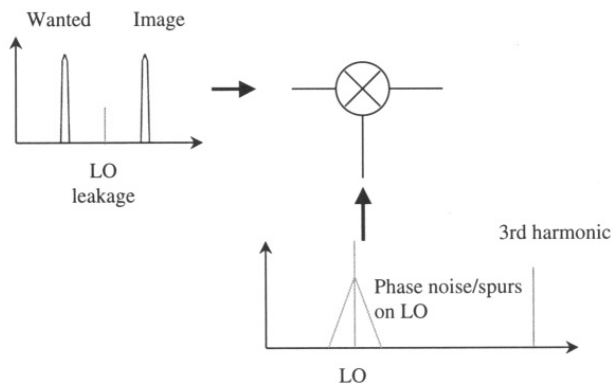


Figure 3.2 Non-idealities in a mixer.

3.1.2 Transmitter concerns

The transmitter must efficiently produce enough output power to ensure that the system has sufficient range whilst ensuring that power is not transmitted into neighbouring channels. It is important to understand how power can be transmitted at unwanted frequencies. Firstly, harmonics will be generated in an efficient power amplifier. These can usually be filtered out before they reach the antenna, but it must be remembered that any filter will also attenuate the wanted output, hence reducing the overall efficiency and increasing the current consumption from the battery. Non-linearity in the power amplifier will spread the spectrum of the signal into neighbouring bands. This behaviour is known as spectral regrowth. It is worse with modulation schemes that are

not constant amplitude. Intermodulation is usually at its worst in the power amplifier itself, but intermodulation throughout the transmitter must be considered.

Broadband noise in the transmitter either due to noise in analogue circuit elements or due to quantisation noise if the signal is generated from a D-to-A converter must be considered. The latter, in particular, will require careful attention to band filtering to ensure low spurious emissions. A low-phase-noise local oscillator is also required to keep the noise emissions to acceptable levels.

In a mechanism similar to phase noise, spurious signals present on the local oscillator will also give rise to spurious emissions on the output signal. Spurious signals can arise from the frequency synthesis process or from unintentional couplings to the VCO. On single-chip systems, the identification of these couplings can be particularly difficult, but attention must be paid to single-ended CMOS input signals which couple strongly to the substrate and also to high-power outputs. An example of the former is given in [1]. These effects are illustrated in Fig. 3.3.

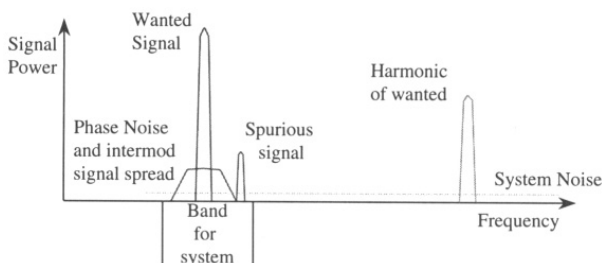


Figure 3.3 Non-idealities in a transmitter.

3.2 RADIO ARCHITECTURES

With the issues discussed in the previous section, the advantages of different architectures can be discussed both in terms of performance and suitability for integration.

3.2.1 Receiver architectures

The double-superheterodyne architecture (Fig. 3.4). This is probably the most commonly employed architecture in current wireless systems. The out-of-band blocking signals are reduced by an RF bandpass filter placed immediately after the antenna. The signal is then amplified by an LNA, which must have a sufficiently low noise to allow detection of weak signals but must also have the dynamic range to handle in-band interferers. The bandpass filter is usually insufficient to reduce signals at the image frequency to the system noise level, and so a second image filter is inserted prior to mixing. To ensure that the image is sufficiently far away from the wanted signal to allow effective filtering, a relatively high first intermediate frequency must be chosen (for 1 to 2 GHz RF systems an IF of 100–200 MHz is common).

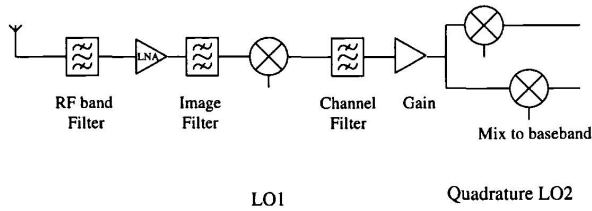


Figure 3.4 The double-superheterodyne architecture.

The mixer must still handle the complete dynamic range of the in-band signal. After the mixer, a SAW filter can be used to achieve the channel filtering. At these frequencies the SAW filter is small but usually has a large in-band loss when complete channel filtering is to be achieved. The output drive of the mixer must therefore boost the signal level to allow for this loss.

Once the interfering channels have been attenuated, the signal can be boosted to a high level (it can be limited if a constant-amplitude modulation scheme is used). The signal is then reduced to baseband frequency for demodulation. It is of course possible to split the channel filtering between the two intermediate frequencies. This will require a greater dynamic range in the second mixer.

This architecture requires the synthesis of two local oscillators, and their frequencies must be chosen so that spurious responses from the radio are kept to a minimum. This aspect of frequency planning, which will not be discussed in more detail here, is a well-understood design process which requires considerable care and experience.

This design requires several external filters and therefore does not lend itself to easy integration as the pin count increases. Moreover, the filters are usually single ended—though this is not essential—and hence achieving isolation between pins becomes an issue. In particular, the channel filter will often need to provide 50 dB of attenuation at key frequencies, thus implying that greater isolation must be achieved between the pins and with respect to signal ground if the filter response is not to be degraded. The image filter can be eliminated if an image-rejecting mixer is used. This will prevent the need to come off chip after the LNA and makes an LNA plus image-reject mixer a useful integrated building block.

The direct-conversion architecture (Fig. 3.5). The direct-conversion receiver, because of its simplicity, appears to offer the best opportunity for integrated systems. Some examples of its use in wireless systems do exist today, but it is not as simple in practice.

Once again, an RF bandpass filter is placed at the input. The LNA's output is passed into the mixer. The LNA must handle the same dynamic range as for the superheterodyne architecture and it must have enough gain to lift weak signals above the noise of the mixer. The mixer however, now converts directly to baseband. Thus the signal is its own image, and channel filtering can now be carried out by low-pass

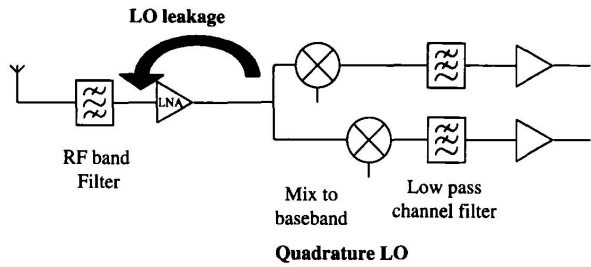


Figure 3.5 The direct-conversion architecture.

baseband filters. Only one local-oscillator frequency needs to be synthesised, and frequency planning is straightforward.

To achieve a precise conversion to baseband, it is important that good quadrature can be achieved at the local-oscillator frequency. Unlike the superheterodyne architecture, this quadrature now needs to be achieved at the RF frequency, which poses some problems that are generally soluble.

The main issue is associated with DC signals which are generated by defects in the mixer. These signals are implicitly in-band, it is therefore difficult, and sometimes impossible, to filter them from the wanted signal. It is necessary to keep them sufficiently below the signal. Unfortunately, the amplification that can be applied to the signal before mixing is limited by the level of in-band interferers which must not overload the mixer.

The DC offsets are caused by a collection of effects. Imbalance in the mixers will lead to a DC output. This is generally a constant quantity and could be cancelled with suitable circuitry. Any leakage of the local oscillator to the input of the receiver will result in a DC signal being generated. If this leakage is via radiation coupling into the antenna, then this may vary with the local environment. Finally, non-linearities in the mixer may cause signals to be generated at DC from other interferers. The latter mechanisms can be time varying, and any offset cancellation needs to be able to respond to time variation. It is only when special precautions are taken to cancel DC offsets that direct-conversion architectures can be used. More details of the direct-conversion issues are given in [2].

Digitisation at IF (Fig. 3.6). This architecture is still a research activity. After the first mixer the full dynamic range of the in-band signals may still exist. This full range needs to be captured by the A-to-D converter. A bandpass sigma-delta converter offers potential for a very flexible system as it does not place high demands on the anti-alias filter. With such a system, the final channel filtering can be made programmable in the DSP filter to allow a single radio architecture to adapt itself to a range of systems. At present the complexity and power consumption in the A-to-D converters provide a significant challenge. As an alternative to bandpass sigma-delta converters, sub-sampled or pipeline converters may provide a viable alternative, but the dynamic range which

can be accommodated will be limited by the number of available bits and the linearity of the converter. It should be noted that a first IF of about 100 MHz will be required to enable adequate image rejection to be obtained at the radio frequency.

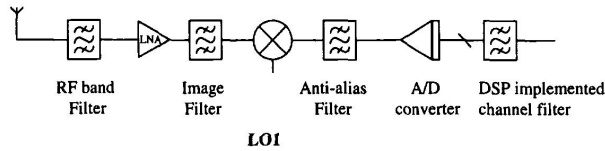


Figure 3.6 Architecture for digitalisation at IF.

If a very-wide-bandwidth A-to-D converter can be used, this system also offers the potential to capture all the activity in the RF band from one receiver. There would be no need to tune to individual channels; this could be done by tuning the digital filter. This would enable radio base-stations to operate on many channels with one radio unit. The number of simultaneous channels that can be received would be limited by the number of DSP filters present.

3.2.2 Transmitter architectures

Direct up-conversion (Fig. 3.7). Direct up-conversion offers a very flexible means of implementing a general modulation. The baseband I and Q signals contain the phase and amplitude modulation information and can be generated digitally. They are converted to analogue signals using D-to-A converters. The major imperfections of the D-to-A converters are quantisation noise and replications of the wanted signal at multiples of the D-to-A clocking frequency. To prevent spurious emissions at unacceptable levels, the signal must be band filtered. The filtering is most easily performed while the signal is at baseband.

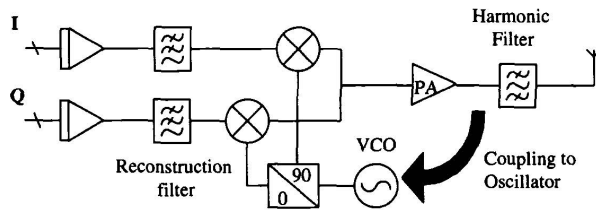


Figure 3.7 Direct up-conversion architecture.

Whilst the signals are at baseband, any non-linearities will create distortion products which will fall in neighbouring channels. It is therefore important that these stages are highly linear as it will be impossible to filter such products once they have been converted to RF.

To convert the signal to RF, a quadrature up-mix is required. The quality of the quadrature will determine the accuracy of the modulation. Any DC offsets will appear as a leakage of the local oscillator through to the output. Once there it will appear as a modulation error. These up-mixer defects must be kept small (about -40 dBc) in order to obtain a high-integrity communication link. Integration and highly balanced circuitry offer this possibility.

The VCO will be very sensitive to pulling from the output signal as it is operating in the centre of the output band. Whether this architecture is implemented in an integrated or a discrete form, unwanted coupling which causes pulling of the VCO will require careful consideration. In integrated form, however, evaluation of the interaction between on-chip components is poorly developed. The design uncertainties in this architecture need to be balanced carefully against its undoubted simplicity when considering integration.

Offset up-conversion (Fig. 3.8). In this architecture, the first up-mix is the quadrature mix and it is to a low first IF. The lower local-oscillator frequency means that a good quadrature local oscillator is more easily built and local oscillator feed-through is less of an issue. As for the direct-conversion architecture, quantisation noise and replications of the signal must be filtered out, though with this architecture the harmonic filter can share this task. The harmonic filter also has the job of removing all the unwanted mixer products from the first mix to ensure they do not appear as unwanted emissions at the output.

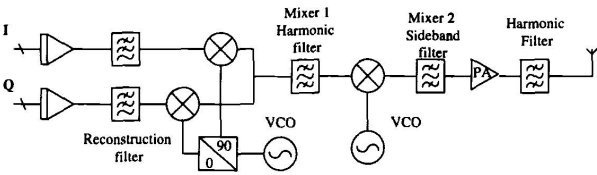


Figure 3.8 Offset up-conversion architecture.

The second mixer needs to be a good single-sideband mixer, as the unwanted sideband will be offset from the wanted RF signal by twice the first intermediate frequency and will cause a significant interference problem if not well suppressed. The final mixer can use a quadrature structure to suppress the unwanted sideband, but even with balanced on-chip components some degree of filtering will be required to keep the unwanted image emission low. To achieve this, the unwanted sideband must lie significantly outside the system band, and this will determine the lowest practical limit for the first intermediate frequency.

The selection of the first and second intermediate frequencies must ensure that unwanted mixer products never result in unacceptably high emissions; this problem is very similar to that of ensuring lack of spurious responses in the double-superhetrodyne receiver. It does however enable an architecture to be designed where the VCOs are operating well away from the power output frequency and are not susceptible to pulling. The need for external filtering will make integration of this type of architec-

ture more difficult, but if an external PA is used anyway, this constraint is not severe. The double up-mixer with external harmonic filter is a very useful integrated block.

Loop-locked and modulated VCO (Fig. 3.9). In this architecture the concept is to reduce the need for signal-path filtering by generating a pure modulated signal from a VCO. With this concept, since the modulation is effectively applied to the VCO input, amplitude modulation of the signal is not possible. However for phase/frequency modulation schemes this is an attractive option.

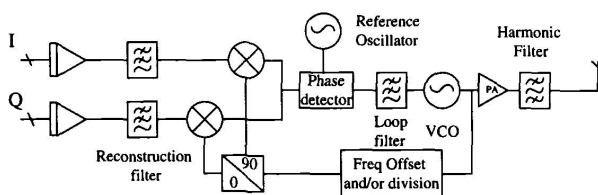


Figure 3.9 Loop-locked and modulated VCO architecture.

To ensure that the modulation is precisely controlled about a desired centre frequency, the modulation is applied in a feedback loop. The aim of the loop is to ensure that phase modulation sampled at the VCO output is in exact anti-phase with the phase at the I and Q baseband inputs. This will then result in a constant-phase signal at the phase detector. Using the high-gain phase detector output to drive the VCO, the resulting loop ensures that this condition is closely maintained. The spurs that can enter the VCO are limited by the bandwidth of the lowpass loop filter, enabling simple control of spurious signals. Because the signal at the PA is very pure, this simplifies the filtering after the PA too. Here, only PA generated harmonics need be removed, which requires a lower-loss output filter and hence improved power efficiency.

The absolute frequency is determined by the reference oscillator and the frequency offset/division block. These need to be carefully planned. Offsetting the frequency requires down-mixing, and it is important to filter unwanted products, as for the superheterodyne receiver. The frequency planning will however ensure that the VCOs should not be susceptible to pulling, and the main VCO is not susceptible as it is phase coherent with the PA. It is not clear, however, whether for a given frequency requirement, this architecture can produce a design with low-frequency external filter which is easily integrable.

3.2.3 Architectures overview

The architectures under consideration are still very similar to those used in radios built from discrete components. SAW filters still play a significant role as they provide good filter characteristics with a very wide dynamic range whilst consuming no power themselves (though we should not overlook the power required to compensate for their loss). Architectures which depend on many SAW filters place limits on the

integrability due to the need to get signals on and off chip, whilst maintaining good signal and ground isolation.

SAWs are also fixed-performance devices, thus if the radio is required to operate across systems, several filter devices which can be switched into the path will be required. Architectures which will allow the programming of the filters will be very attractive in the future for such multi-mode systems. Architectures where the signal is digitised with a wide dynamic range and is filtered and demodulated using the power and low cost of data processing implemented in a small-dimension CMOS process look attractive for this purpose. However, sufficiently high performance and sufficiently low power consumption has not been achieved in the analogue to digital conversion.

Implementation in integrated form offers significant advantages for matched and balanced circuitry. Thus architectures which exploit this, for example image-rejecting mixer based circuits, will evolve as prime candidates for integrated radios. However, in any integrated radio, we need to be able to guarantee isolation between the sections. Currently, isolation is achieved by PCB and screened-compartment design. The move to integration will require that isolation is achieved on chip. At present, design skills in this area are low. The choice of architecture to minimise isolation requirements, and the development of techniques to calculate the isolation across the chip are going to be key factors, if the level of integration of radios is to increase.

3.3 SOME DESIGN ISSUES

In this section, some of the design issues for integrating radio circuitry in silicon technology are introduced.

3.3.1 Power supply and ground coupling

The essential problem here is that RF currents flowing in the power-supply and ground leads can drop significant potentials across track and bond inductances. Moreover, capacitances can resonate with the bond/track inductances causing the impedance seen by the circuitry to become high at certain frequencies. These frequencies must be separated from key circuit frequencies. Supply return currents may cause potentials to be dropped on signal grounds. This can give rise to feedback to the input, poor isolation and even instability. Fig. 3.10 shows the situation for power supply decoupling.

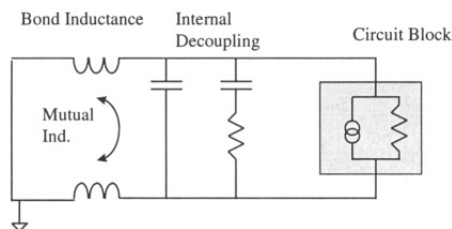


Figure 3.10 Power supply decoupling.

In this circuit, some on-chip capacitance has been added to decouple the power supply. If we assume the power supply is perfectly decoupled close to the chip, then the bond inductance—modified by any mutual coupling between the supply and ground lines—will resonate with the internal decoupling capacitance. If the on-chip circuitry takes a significant current at this resonant frequency (represented by the current source) then a large power supply ripple will occur. This ripple is limited by the loss in the circuitry itself and the loss in the decoupling capacitors. Simulation can be used to estimate the magnitude of this ripple and also to decide on the values of the capacitance to use (to put the resonance in an innocuous range) and how much damping to provide if the circuit does not provide sufficient loss.

As general design principles:

- Separate power supplies for blocks which require isolation.
- Keep ground and power supply pins close to exploit mutual inductance.
- Add on-chip decoupling, choose resonance away from circuit frequencies.
- Damp resonance if circuit does not provide sufficient loss.
- Try and separate signal grounds and ground power returns if possible.

3.3.2 Substrate coupling

As a general rule, as interconnection and circuitry scales down, cross-talk capacitance decreases. This occurs because the length of interconnect required decreases as the circuit compacts with scaling, whilst the cross-talk capacitance per unit length varies relatively little. This implies that to make higher speed circuitry, all dimensions will need to be reduced, and integration is a natural route forward.

However, with integration, there is also the possibility that signals will couple through the substrate. With silicon technology, where the substrate is conducting, it is possible for the signal to be conducted through the substrate. Fig. 3.11 illustrates this.

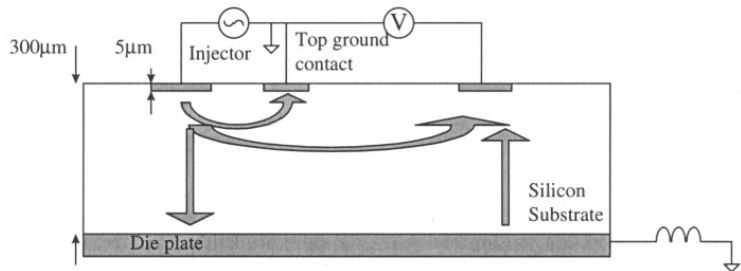


Figure 3.11 Coupling through the substrate.

Devices sitting in the top few microns of the silicon can inject charge capacitively into the substrate. Even on a silicon-on-insulator (SOI) technology this will happen,

but a good SOI technology will have a low coupling capacitance. The major coupling will be to nearby devices or local contacts. Some of the charge will be injected more deeply and couple to distant circuit elements. This will be a source of interference. Finally, some charge will return via the rear die plate. The die plate potential will vary in response to the charge and the return path impedance (which will include capacitive coupling to other parts of the circuit and the impedance to ground of the die plate). If the die plate does not have a low ground impedance then it can couple the injected signal effectively across the whole chip.

These effects are being studied now. Means of simulation are being developed (for example, see [3]). As a general rule, balanced signals will always help injection, and local grounded substrate contacts are beneficial.

3.3.3 *On-chip oscillators*

Oscillators on chip always need careful consideration. Although the maintaining amplifier can be easily accommodated on chip, there is an advantage to keeping the whole oscillator on chip. This however requires an on-chip resonator. The phase-noise performance is directly dependent on the resonator's quality factor. Spiral inductors can be manufactured on chip, but their quality factor is limited by the resistance of the metal used to wind the inductor (losses in the substrate are usually a smaller effect). Measurements of the phase noise achieved using a 2 μm -thick Al/Si/Cu metallisation show that whilst the phase noise required for DECT is attainable, the requirements of a system with much narrower bands such as GSM is rather too demanding.

The second issue to be considered with on-chip oscillators is pulling. The oscillator can be considered as a highly shaped and amplified noise source. Essentially, at the resonant frequency, the feedback circuit provides just enough gain to amplify the natural circuit noise to the level required, so this gain is controlled by the limiting mechanism. Away from the natural resonant frequency, the loop gain falls away; the higher the Q the more rapidly the gain falls off. This leads to two defect mechanisms: if a large signal close to resonance couples into the resonator, it will get amplified until it causes the limiter to reduce the gain. The noise content at the output will fall away, and the oscillator will have locked to the interferer. If an interferer is coupled in at a much lower level, then it will not trigger the limiting action; the shaped noise will still exist, but the spurious tone will interact with the oscillations and modulate the frequency. A side-tone will appear at the interferer frequency and a second one on the other side of the main tone—characteristic of FM modulation. Now the sensitivity to such an interferer depends on how close to the main tone the interferer is. If it is at the resonant frequency itself, the coupling only needs to be at circuit noise levels, the level required increases as the coupling tone is offset in frequency. The higher the resonator's Q the higher the level of the interfering signal that can be tolerated. Since very small signals may pull an oscillator, the choice of architecture is critical. The aim, if an on-chip oscillator is to be included, will always be to ensure that spurious signals which could pull or lock the VCO are not present.

References

- [1] Jeff Durec, “An Integrated Silicon Bipolar Receiver Subsystem for 900 MHz ISM Band Application”, *JSSCC*, vol. 33, Sept. 1998, pp. 1352–1372.
- [2] Behzad Razavi, “Design Considerations for Direct Conversion Receivers”, *IEEE Transaction on Circuits and Systems-II*, Vol. 44, June 1997, pp. 428–435.
- [3] Ranjit Gharpurey and Robert G. Meyer, “Modeling and Analysis of Substrate Coupling in Integrated Circuits”, *IEEE JSSC*, vol. 31, Mar. 1996, pp. 344–353.