Direct-Conversion Radio Transceivers for Digital Communications

Asad A. Abidi, *Senior Member, IEEE*

***Abstract*—Direct-conversion is an alternative wireless receiver architecture to the well-established superheterodyne, particularly for highly integrated, low-power terminals. Its fundamental advantage is that the received signal is amplified and filtered at baseband rather than at some high intermediate frequency. This means lower current drain in the amplifiers and active filters and a simpler task of image-rejection. There is considerable interest to use it in digital cellular telephones and miniature radio messaging systems. This paper briefly covers case studies in the use of direct-conversion receivers and transmitters and summarizes some of the key problems in their implementations. Solutions to these problems arise not only from more appropriate circuit design but also from exploiting system characteristics, such as the modulation format in the system. Baseband digital signal processing must be coupled to the analog front-end to make direct-conversion transceivers a practical reality.**

I. INTRODUCTION

THE CURRENT interest in portable wireless communications devices is prompting research into new IC technologies, circuit configurations and transceiver architectures. Low-power miniature radio transceivers are sought to communicate digital data in cellular telephones, wireless networks, and radio messaging systems. While transistor *technology scaling* and *improved circuit techniques* will contribute evolutionary advances towards this goal, *architectural innovations* in the transceiver may lead to revolutionary improvements [1]. It is in this context that there is a resurgence of interest in direct-conversion.

The superheterodyne receiver, which Armstrong introduced in 1918 [2], is generally thought to be the receiver of choice owing to its high selectivity and sensitivity. Something like 98% of radio receivers use this architecture. In a superheterodyne receiver, the input signal is first amplified at RF in a tuned stage, then converted by an offset-frequency local oscillator to a lower intermediate frequency (IF), and substantially amplified in a tuned IF "strip" containing highly-selective passive bandpass filters. The role of the various filters is illustrated by the typical frequency plan of a superheterodyne receiver (Fig. 1). The IF must be sufficiently high so that the *image channel* lies in the stopband of the RF preselection filter or the antenna, otherwise the IF filter will pass this channel unattenuated in its own image passband. These considerations determine the familiar intermediate frequencies used in radio and TV receivers.

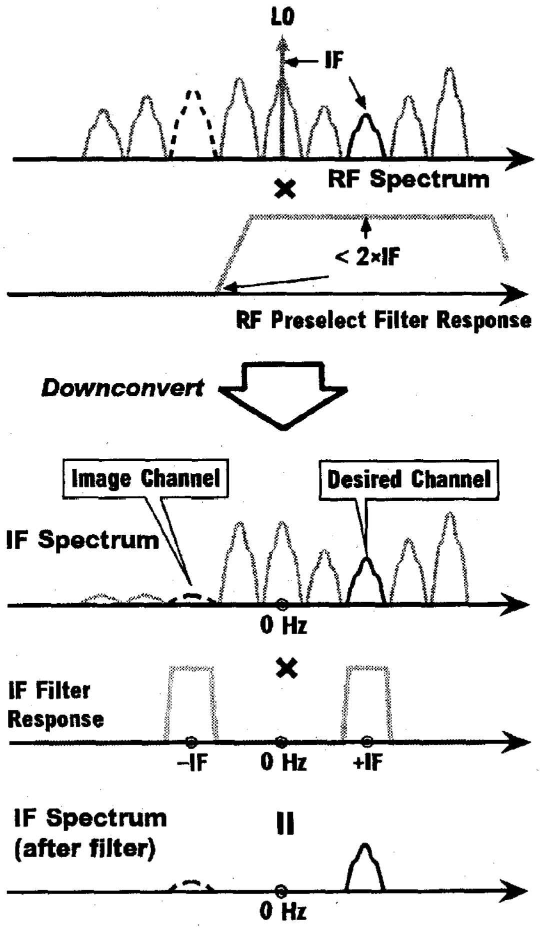


Fig. 1. Frequency plan of a superheterodyne receiver. Choice of IF is governed by width of preselect filter passband. The RF preselect and IF filters work together to select the desired channel.

In a standard broadcast FM receiver, for instance, the 10.7 MHz IF guarantees that the image channel lies outside the 20 MHz wide FM band. Therefore, even if the preselect filter inadequately suppresses the image, which is assumed not to be an FM signal, the subsequent frequency-discriminating detector will inherently tend to reject it. This relaxes the selectivity of the 100 MHz preselect filter, which may be constructed with either a single or ganged collection of *LC*-tuned circuits. Ceramic filters are an enabling technology for the 10.7 MHz FM IF. These small and cheap filters

Manuscript received May 22, 1995; revised August 29, 1995.

The author is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA.

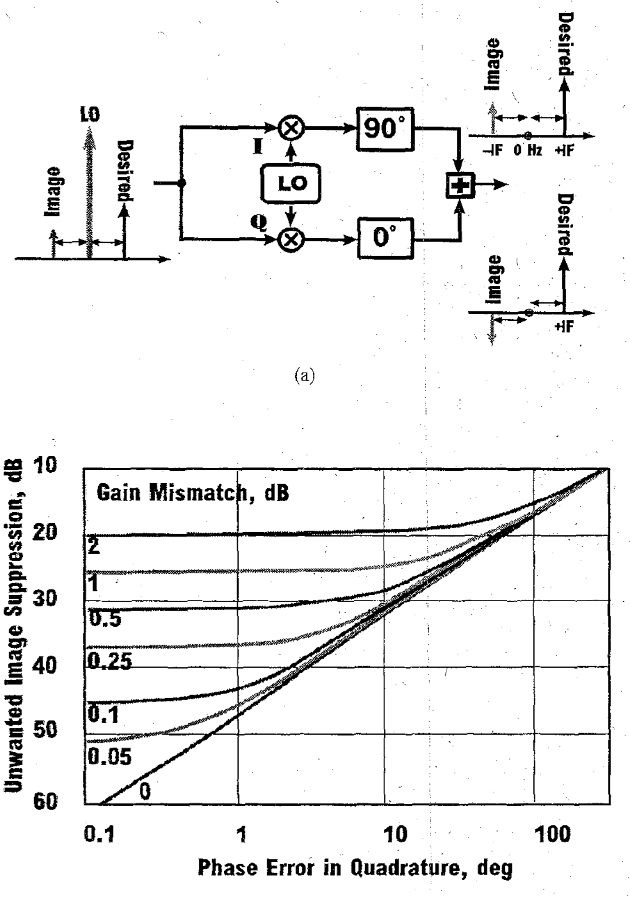
IEEE Log Number 9415818.

0018–9200/95$04.00

offer a narrow passband and good stopband attenuation, and they are widely used. The traditional 43.5 MHz IF in a TV receiver cannot suppress the image across the entire VHF, hyperband, and UHF bands, so as the receiver is tuned across various sub-bands, one of an array of RF narrowband filters is switched into the RF front-end [3] to suppress the image channel. Many analog cellular telephones use a 90 MHz IF. Amplification and filtering at these high intermediate frequencies between 10–100 MHz comes at the price of *power dissipation* because transistors must be biased at large currents to drive the parasitics and the low characteristic impedance of the passive IF filters. Further, the IF strip may require a large number of *off-chip passive* components, which add to receiver size. Although these are not serious problems for tabletop receivers—the easy alignment of a superheterodyne, resulting in a high selectivity, was always one of its strengths—they may become limitations in miniature, low power transceivers.

Wireless receivers must often handle very weak channels existing side-by-side with very strong channels in the same band. Thus, in addition to a minimum *stopband attenuation* to suppress the image channel, the filter must also have a wide *dynamic range*, that is, the ability to handle strong signals without distortion while remaining sensitive to weak signals above the intrinsic noise level in the passband. In this respect, passive filters are almost always superior, as the small-signal handling of active bandpass filters is limited by a fundamentally higher noise level [4], and nonlinearity in the active device tends to distort large signals. The dynamic range of active filters may only be increased at the expense of capacitor size and power dissipation.

Although most often a passive RF preselect filter attenuates the image channel, it may also be suppressed by selective *signal cancellation*. Here, the entire RF spectrum is downconverted to an IF in two identical mixers driven by quadrature phases of a local oscillator (LO). The downconverted spectra in the two branches are subjected to a 90° phase-shift relative to one another and then added (Fig. 2(a)). With appropriate design, this arrangement downconverts the desired channel to IF with the same phase in the two branches and the image channel to -IF but antiphase in the two branches. After addition, the desired channel appears at the output with double strength, while the image channel subtracts and disappears. This *image-reject downconverter* is the dual of Weaver's celebrated phasing method of sideband selection [5], which is discussed later in the section on transmitters. In practice, departures from quadrature in the two LO signals and gain mismatch in the two branches will limit the extent of signal cancellation (Fig. 2(b)). When used in a receiver, the effectiveness of this image-suppression method is further limited by the wide dynamic range of radio signals. If the unwanted image channel is much stronger than the desired one, then after imperfect signal cancellation, it may only be suppressed to a comparable level to the desired channel, resulting in an intolerably large interference. Nevertheless, this type of downconverter is used, for instance, in a single-chip broadcast FM receiver with a low IF of 150 kHz [6].



(a)

(b)

Fig. 2. (a) The image-canceling downconversion mixer. The desired signal appears in-phase in the two branches and the undesired signal anti-phase. Allpass filters may be used to synthesize 90° phase shift. (b) Unwanted signal suppression as a function of errors in phase from ideal quadrature, with gain mismatch in the two branches as a parameter.

II. THE DIRECT-CONVERSION ARCHITECTURE

Suppose that the IF in a superheterodyne is reduced to zero. The LO will then translate the center of the desired channel to 0 Hz, and the portion of the channel translated to the negative frequency half-axis becomes the image to the other half of the *same* channel translated to the positive frequency half-axis (Fig. 3). The downconverted signal must be reconstituted by a phasing method of the type described above, otherwise the negative-frequency half-channel will fold over and superpose on to the positive-frequency half-channel. Zero-IF, therefore, mandates quadrature downconversion into two arms and a vector-detection scheme. However, this scheme does not suffer from the strong-image problem when the image-reject downconverter is used in a nonzero IF heterodyne receiver, and the typical gain mismatches and phase errors in the two branches cause only a small loss in detected SNR. A lowpass filter, which is in effect a bandpass centered at dc when the negative frequency axis is included, may be used to select the desired channel and to reject all adjacent channels. Therefore, RF preselection may in principle be

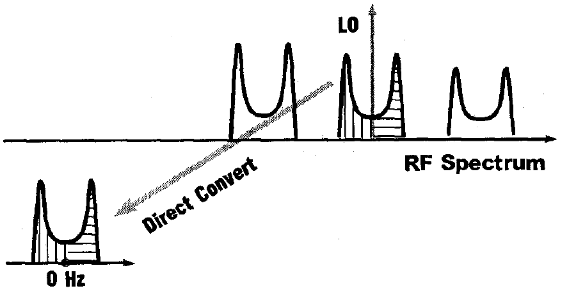


Fig. 3. Spectrum before and after direct-conversion.

eliminated because there is no image channel. In practice, it is still required to suppress strong out-of-band signals that may create large intermodulation distortion in the front-end prior to baseband channel selection and to avoid harmonic downconversion. There is also the advantage that if a high-order active filter is used for channel selection, it will dissipate lower power and occupy a smaller chip area at a given dynamic range than an active bandpass filter with the same selectivity centered at a high IF [4]. All amplification past the front-end is also at baseband, and therefore consumes a small power. This zero-IF scheme is also called *direct-conversion*. When the local oscillator is synchronized in phase with the incoming carrier frequency, the receiver is called a *homodyne*.

As early as 1924, radio pioneers had considered use of homodyne architectures for single vacuum-tube receivers, but it was a homodyne measuring instrument for carrier-based telephony built in 1947 that first employed a high-order lowpass filter for channel-selection [7]. Thereafter, the concept lay dormant, until it was revived in 1980 in the radio-paging receiver, the first miniature digital wireless device for personal communication to attain widespread consumer use.

III. DIRECT-CONVERSION FSK RECEIVERS

Digital data in broadcast paging modulates the carrier by frequency-shift keying (FSK). The carrier frequencies may lie in the 400 MHz or the 900 MHz bands and binary data at 512 b/s or 1.2 kb/s rates shifts the carrier frequency by ±4.5 kHz. This large modulation index results in a spectrum with two lobes symmetrically offset around the carrier (Fig. 3). Vance at ITT Standard Telecommunications Labs was the first to apply direct-conversion to this signal spectrum with a *single-chip* paging receiver [8], thus establishing a key concept for small, light paging receivers. Not all pagers today, though, use direct-conversion; some continue to use the superheterodyne implementation for higher performance [9].

Following a single-stage of RF amplification in this simple FSK receiver (Fig. 4), a local oscillator (LO) tuned to the incoming carrier downconverts the center of the desired paging channel to dc. In fact, quadrature phases of the LO downconvert the signal into two branches, labeled I and Q, enabling the detector to discriminate the signal at positive and negative frequencies (i.e., data 1's and 0's). A high-order *lowpass filter* in each branch with a cut-off at about 10 kHz selects the desired channel, while all other channels fall into the filter stopband. This may be integrated as a low-power active filter. A single-chip paging receiver from Philips uses a tenth-order continuous-time lowpass filter [10]. The data is encoded in the zero-crossings of the downconverted and filtered signal, which a limiter then amplifies to logic levels, eliminating the need for AGC. However, dc offsets directly add to the downconverted signal and may be so large as to disable zero-crossings in the limiter output, causing the receiver to fail to detect data. This problem is overcome by capacitively coupling the baseband signal path into the limiter to null these offsets. Some of the consequences of capacitive coupling are discussed in Section VI.

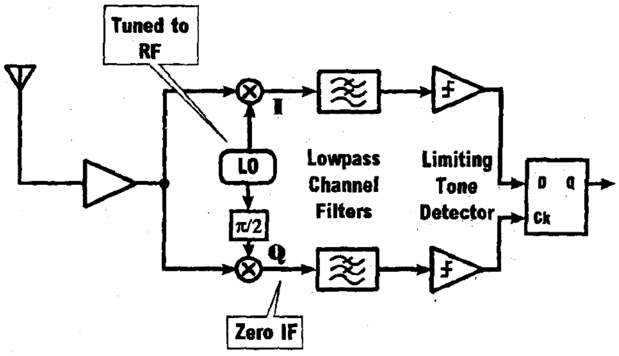


Fig. 4. Block diagram of a direct-conversion FSK receiver as may be used for radiopaging signals.

The detector is, in principle, only a flip-flop, driven at the D input by the I branch limiter and at the CK input by the Q branch limiter. The flip-flop output attains one steady-state if transitions at the CK input lead the transitions at the D-input and the other state if they lag. This corresponds exactly to a positive or negative frequency shift of the carrier, and thus, the data. Although this simple detector is found in many FSK receivers, it is susceptible to upset from a single noise impulse at the input. A more sophisticated detector oversamples the limiter output at a multiple of the data rate, thereby more finely quantizing the zero-crossing instant, and correlates this with quadrature phases of the expected frequency shift (4.5 kHz in the paging channel). The correlated output from the I and Q channels is integrated over a bit period, and the bit decision is made depending on which of the integrators first crosses a preset threshold. Correlation reduces the noise bandwidth. A one-bit implementation of this correlation detector in a spread-spectrum FSK receiver shows that it is very compact and dissipates a small power [11].

There are now many low-power, single-chip bipolar IC's implementing direct-conversion paging receivers [10], some operating at a supply as low as 1.1 V [12]. The on-chip capacitors required by the two active lowpass filters and ac coupling after downconversion occupy a large portion of the total die area. Aside from a reference crystal, the circuits only need some miniature off-chip inductors for the tuned RF amplifier loads and sometimes for the quadrature phase-shift network. The complete pager, including the microprocessor

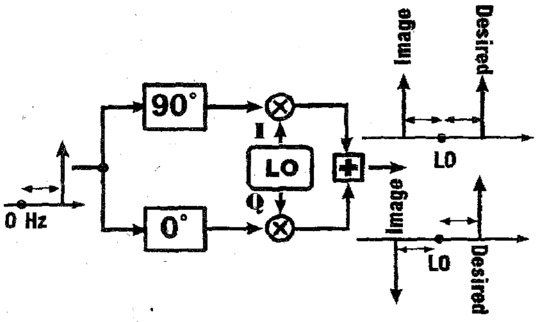


Fig. 5. A direct-upconversion mixer using the phasing method for selecting one sideband. Either one of the sidebands may be selected by combining the branches with the appropriate sign.

and display driver, may be a two-chip device with battery life in excess of six months.

One may appreciate the simplicity of a direct-conversion state-of-the-art paging receiver by comparing its inventory of parts with a superheterodyne implementation of the same built in a comparable technology [13]. The superheterodyne requires one more crystal, two trim capacitors, and a SAW filter, which together add a significant fraction to the total parts count, thus increasing the physical volume of the receiver and its power dissipation.

IV. DIRECT-CONVERSION SINGLE-SIDEBAND SYNTHESIZERS

For reasons of spectral efficiency, the transmitted signal in digital communications is usually single-sideband with suppressed-carrier. It would require an RF filter with a very sharp transition band to suppress one sideband on the modulated carrier while passing the other. The much more practical solution is the phasing method [5], where the modulated signal is first synthesized in quadrature at baseband, *directly-upconverted* into two branches by a quadrature LO centered at the carrier frequency, and added or subtracted to select either the upper or lower sideband (Fig. 5). The phasing method of sideband selection has been used for many years in single-sideband communication transceivers.

The unwanted sideband is suppressed to an extent limited by the gain mismatch in the two upconversion branches and by departures from quadrature in the two LO outputs (Fig. 2(b)). The dc offsets in the branches produce an output tone at the LO frequency. The unwanted sideband and LO leakage are unavoidable spurious emissions in the transmitted spectrum. Although the two upconversion branches will match well on the same IC, a gain mismatch as small as 1% (0.1 dB) limits unwanted sideband suppression to about 45 dB. With this gain mismatch, a phase-error of up to 1° is tolerable between the two LO outputs before the unwanted sideband grows further in relative amplitude. These mismatches may be trimmed at time of transceiver manufacture or self-calibrated with loopback modes that are activated during idle times to sense and suppress the unwanted sideband. Some trimming and adaptive methods are discussed in Section VII.

As the LO frequency in a direct-upconverter is centered in the transmit band, energy at this frequency may be spuriously radiated through parasitic unbalanced coupling into the power amplifier or antenna. For instance, the single-ended signal produced by an on-chip oscillator circuit tuned with an off-chip resonator may couple to the power amplifier input across pins of the RF package. Frequency-offset multi-step upconversion schemes, which are the dual of a heterodyne downconverter, have been proposed [14] to combat this coupling problem. However, as LO phase-noise in a transmitter appears as noise added to the emitted signal, a process called *reciprocal mixing* in the radio literature, direct upconversion has the advantage over a frequency-offset scheme that only *one* LO contributes noise. Other spurious output tones in a single-sideband transmitter may arise from parasitic remixing of the modulated output with the baseband signal and by intermodulation distortion in the output stage [15]. Balanced circuit topologies, on-chip LO's that require no external resonators [1], [16], and the lowered transmit power levels required in microcells, are all expected to lessen the magnitude of these problems.

V. DIRECT-CONVERSION RECEIVERS FOR DIGITAL CELLULAR TELEPHONES

Designers of portable digital cellular telephones are very interested in low-power radio architectures. Several integrated receiver and transmitter IC's conforming to established standards such as GSM and DECT have been developed in the past few years. This section summarizes some of their main features.

All transmitters in these portable phones use direct upconversion to produce a single-sideband output. In receivers, however, the superheterodyne architecture is more common. For instance, a 900 MHz bipolar IC GSM receiver from Siemens [17] downconverts the amplified RF signal from an off-chip low-noise amplifier to an IF of 45–90 MHz (Fig. 6). At this IF, the image lies in the stopband of the fixed RF preselect SAW filter. The amplified IF signal is sent to another off-chip SAW filter to reject adjacent channels. A quadrature mixer then downconverts the signal to baseband, and the vector baseband signal is finally detected. This architecture is preserved in later generations of this transceiver operating up to 2 GHz for DECT use [18], [19]. Other recent GSM transceivers build on a similar single superheterodyne architecture [20], in one case with a very high IF of 400 MHz [21]. Alcatel has publicized its use of direct-conversion in GSM and DECT receivers [22]–[24], although others [25] are exploring its possible use, and not all companies using direct conversion have published their experience. Alcatel's RF front-end is a relatively small silicon bipolar IC (Fig. 7), and the remainder of the signal processing, including lowpass channel-select filtering, takes place at baseband in a mixed analog-digital CMOS IC [26].

Given the many decades of familiarity with the superheterodyne, there will likely remain some reluctance towards adopting a new architecture until there is widespread experience in its effectiveness. However, direct-conversion also suffers from some unique problems to which the superheterodyne is immune. These are discussed in the following section.

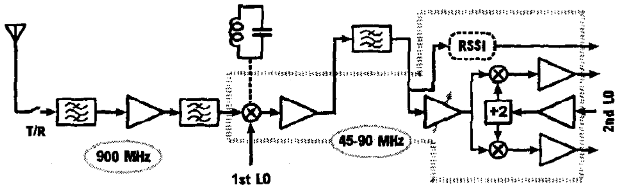


Fig. 6. A superheterodyne receiver for a digital cellular telephone.

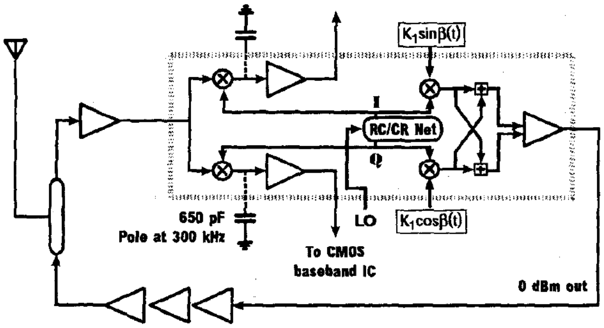
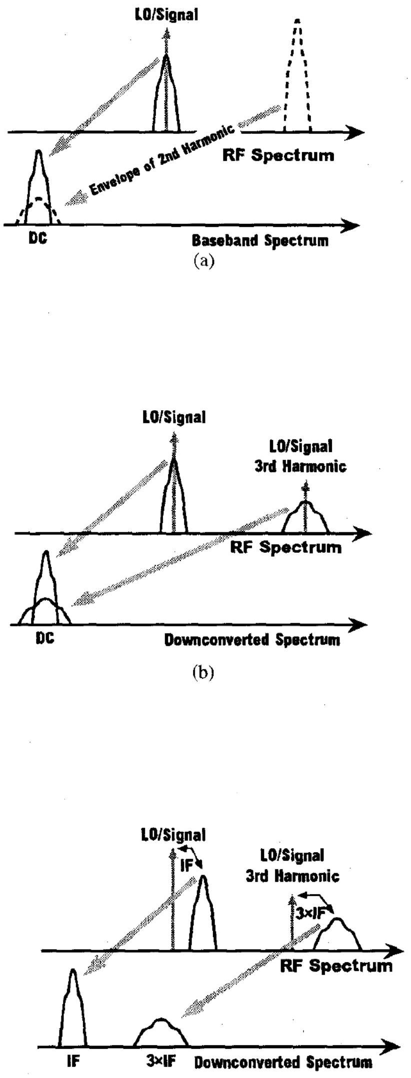


Fig. 7. A direct-conversion receiver and transmitter for a digital cellular telephone.

VI. Problems in Direct-Conversion Receivers

Among the problems in direct-conversion receivers, *spurious LO leakage* is probably best known. This arises because the LO in a direct-conversion receiver is tuned exactly to the center of the LNA and antenna passbands. In receive mode, a small fraction of the LO energy may make its way back to the antenna through the mixer and LNA, owing to their finite reverse isolation, or couple into the antenna through external leads, and then radiate out [27]. This becomes an in-band interferer to other nearby receivers tuned to the same band, and for some of them it may even be stronger than the desired signal. Regulatory bodies such as the FCC strictly limit the magnitude of this type of spurious LO emission. The problem is much less severe in a superheterodyne whose LO frequency usually lies outside the antenna passband. However, experimental studies [28] suggest that standard shielding in the receiver may control LO leakage to the point that it does not seriously handicap the use of direct-conversion.

Distortion produced by strong signals in the downconversion mixer will cause the sensitivity of a direct-conversion receiver to degrade more rapidly than of the superheterodyne. Second-order distortion in a single-ended mixer will rectify the envelope of an amplitude-modulated RF input such as QPSK data to produce spurious baseband spectral energy centered at dc, which then adds to the desired downconverted signal [25], [27] (Fig. 8(a)). This is particularly serious if the envelope is that of a large unwanted signal lying in the preselect filter passband, which has not yet been rejected by the baseband channel-select filter. The most effective solution is to use *balanced circuits* in the RF front-end, particularly the mixer, which will only create odd-order distortion.[1]



(a)

(b)

(c)

Fig. 8. Spurious downconversion caused by nonlinearity in a direct-conversion receiver. (a) Second-order distortion detects envelope of near-band interferer at baseband, overlaying desired signal. (b) LO harmonics downconvert signal harmonics to baseband, resulting in interference, (c) whereas in a superheterodyne downconverted harmonic products lie in IF stopband.

However, even in a balanced circuit, the third harmonic of the desired signal may downconvert the third LO overtone to create spurious dc energy competing with the fundamental downconverted signal (Fig. 8(b)), whereas in a superheterodyne this downconverted component lies in the stopband of the IF filter (Fig. 8(c)). This is a small effect to the extent

---

[1] There is no fundamental loss of noise figure in a balanced front-end. When the antenna signal drives a balanced low-noise amplifier through a power-splitting balun, the noise figure is exactly the same as directly driving the signal into a single-ended half circuit. However, the balanced circuit drains twice the current of the single-ended half circuit.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.