IQ-imbalance and its compensation for non-ideal analog receivers comprising frequency-selective components

M. Mailand, R. Richter, and H.-J. Jentschel
Dresden University of Technology, Institute of Traffic Information Systems, 01062 Dresden, Germany

Abstract. Within current implementations of mobile terminals, more and more analog components are replaced by appropriate digital processing. On the one hand, the analog front-ends become less complex. On the other hand, more digital signal processing is required to compensate for the spurious effects of the front-end. In this article, the frequency-selective imbalance of the in-phase and quadrature-phase signals is addressed. A closed representation of arbitrary signals being processed by an arbitrary imbalanced analog front-end is provided. The analysis is valid for both, direct conversion and intermediate frequency (IF) reception. With the consideration of practical variations of amplitude and phase impairments, the influence of only the frequency-dependent portions of the impairments is investigated. It is shown, that the compensation of the quasi-linear impairments is sufficient and complex deconvolutive IQ-regeneration procedures are not stringently required to obtain sufficient signal qualities.

1 Introduction

In modern mobile communication receivers, more and more analog components are replaced by digital building blocks. One the one hand, this leads to less circuitry effort (chip size, power consumption, etc.) in the analog domain. On the other hand, additional digital signal processing is required in order to remove spurious effect of the sub-optimum analog front-end.

In this article, we investigate a certain type of analog signal distortions, i.e. the in-phase (I) and quadrature-phase (Q) imbalance. For heterodyne or intermediate frequency (IF) reception, IQ-imbalance causes the well-known image problem. There, the respective image channel mixes partially onto the wanted signal during the IF-down conversion, (Valkama et al., 2001). Within an homodyne, direct conversion or zero-IF reception, IQ-imbalance leads to a distortion of the IQ-signals themselves within the respective wanted channel. For both frequency down conversion concepts, the IQ-imbalance is a serious issue degrading the reception performance. This spurious effect occurs mainly due to amplitude- and phase-impairments between the local oscillator paths as well as due to mismatches between the respective IQ-branches after the analog down conversion.

For the case of wideband transmission, the anyway distorting IQ-imbalance mixing becomes an IQ-imbalance convolution, due to the frequency selectivity of the analog front-end components after the down conversion. Thereby, especially different impulse responses of filters and IF- or baseband-amplifiers have significant influence.

For direct conversion and IF-conversion, we will show for transmissions utilizing wideband signals, that the impact of the IQ-imbalance convolution or the frequency selectivity of the components, respectively, has not to be considered stringently. Mainly, the mean values of the phase- and amplitude impairments of the IQ-imbalance convolution cause most of the signal distortions. Therefore, it will be sufficient to compensate for the frequency-selective IQ-imbalance convolution with methods primarily developed for ‘simple’ IQ-imbalance compensation, e.g. by utilizing blind source separation (BSS) algorithms, (Cardoso et al., 1996).

2 Receiver front-ends

The technique of direct down conversion (DDC) transforms the RF-signal directly down to baseband. For that purpose, the LO is set to the carrier frequency \( \omega_{LO} = \omega_{RF} \) of the wanted channel. Due to temperature dependencies, production imperfections etc., the analog components in the I- and Q-path can not be perfectly matched. The amplitude as well
as the actual phase of the respective LO-signal to the I- or Q-path will differ from the respective optimal values resulting in phase φ_i and amplitude g_i imbalances. Therefore, the ideal LO-signal changes into

\[ \xi_{LO}(t) = g_i \cdot \cos(\omega_{LO} t + \phi_i) - j \cdot g_i \cdot \sin(\omega_{LO} t + \phi_i) \]  

for the distorted paths (see Fig. 1).

Independent of the respective front-end architecture, there will be different signal paths for the I-signal and the Q-signal. These signal paths are also expected to be unmatched in the realistic case. Hence, they influence the IQ-components by an additional scaling, i.e. path mismatch \( M_i \) for the \( i \)th path.

For further investigations, we assume that any type of receiver being regarded comprises some direct current (DC) offset cancellation.

The influence of flicker noise or 1/f-noise which is an additional serious problem especially for CMOS implementations of DDC receivers is not addressed in this paper.

2.1 Multiplicative mixing

Multiplicative mixing is the conventional implementation of direct down conversion. Here, the RF-signal is multiplied by the original LO-signal within one path (in-phase signal, I) and by a 90° rotated LO-signal in a second path (quadrature-phase signal, Q), i.e. in principle \( x(t) = \text{LPF}(s_{RF}(t) \cdot s_{LO}(t)) \). The upper signal branch is assumed not to be matched to the second path. This leads to a relative path mismatch \( M_i \). In Fig. 1a, the receiver front-end with the respective impairments is shown. After mixing, the low pass filter (LPF) suppresses the unwanted higher frequency terms. Thus, almost only the IQ-signals of interest will be digitized. According to Fig. 1a and considering all the impairments, we get:

\[ z'_I(k) = \frac{1}{2} g_1 M_1 \cdot \left( s_I(k) \cos(\phi_1) + s_Q(k) \sin(\phi_1) \right) \]  

\[ z'_Q(k) = \frac{1}{2} g_2 \cdot \left( -s_I(k) \sin(\phi_2) + s_Q(k) \cos(\phi_2) \right) \]  

In general, the Eqs. (2) and (3) can be rewritten in a matrix form to consider all possible, linear imbalances (except DC-offsets):

\[ z' = As = \begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \end{bmatrix} \begin{bmatrix} s_I(k) \\ s_Q(k) \end{bmatrix} \]  

We consider to have arbitrary gain and phase impairments within each of the IQ-paths, although a relative IQ-imbalance was sufficient. (Valkama et al., 2001). Hereupon, we define the signal-to-interference ratio (SIR) for the IQ-branches:

\[ \text{SIR}_I = 20 \cdot \log_{10} \left( \frac{|A_1|}{|B_1|} \right), \text{ SIR}_Q = 20 \cdot \log_{10} \left( \frac{|B_2|}{|A_2|} \right) \]  

2.2 Additive mixing

Direct down conversion by additive mixing gained a lot of attention in the recent years. Usually, the six-port technology is applied. In the use as communications receiver the former six-port was reduced to a five-port structure. Figure 1b shows the general architecture for additive DDC utilizing a five-port. Clearly, the mixers (multiplicative elements in Fig. 1a) are replaced by appropriate summing elements followed by a square-law device (in general, a simple nonlinear amplification component, e.g. a diode). The advantages compared to the mixer-based concept are e.g., better robustness for RF-signal power level fluctuations, DC-offsets can be cancelled more easily and it is possible to implement broadband receivers in a comparably simple manner, (Ratni et al., 2002).

The down converted signal before analog-to-digital conversion (A/D) of the front-end can generally be denoted with

\[ z'_I(t) = s_I(t) g_i \cdot \cos(\phi_i) + s_Q(t) g_i \cdot \sin(\phi_i) \]  

\[ + 2 \cdot M_i \cdot (s_I(t) g_i \cdot \cos(\phi_i) + s_Q(t) g_i \cdot \sin(\phi_i)) \]  

The phase shift elements \( \phi_i \) and the attenuation/amplification components \( g_i \) within the five-port comprise known, wanted values as well as impairments. Hence, they contribute to the indeterminacy of the front-end transmission due to spurious effects.

The rectified wave which comprises the DC-offsets can be easily removed by a simple front-end calibration (Hentschel, 2004) or by the implementation of an appropriate front-end architecture (Mailand et al., WCNC 2005).

Therefore, we obtain measurements which consist only of linear mixtures of the IQ-signals, i.e.

\[ z'_I(k) = 2 \cdot M_i \cdot (s_I(k) g_i \cdot \cos(\phi_i) + s_Q(k) g_i \cdot \sin(\phi_i)) \] in
Eq. (6). In consequence, the transmission of the DDC front-end with additive mixing is determined by

\[ z' = A \mathbf{s} = \begin{bmatrix} z'_I(k) \\ z'_Q(k) \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \end{bmatrix} \begin{bmatrix} s_I(k) \\ s_Q(k) \end{bmatrix}. \]  

(7)

The phase parameters \( \varphi_i \in [0, 2\pi) \) have to be different of each other, to assure a possible separation of \( s_I(t) \) and \( s_Q(t) \), i.e. to guarantee a nonsingular mixing matrix \( \mathbf{A} \).

The composition of the baseband signals or front-end measurements \( x_i(k) \) is systematically a mixture of I-signals and Q-signals for additive mixing. Utilizing multiplicative mixing, the IQ-mixture within one measurement \( z'_i(k) \) is due to front-end impairments.

Nevertheless, although the analog processing is significantly different, the general description of direct down conversion transmission, considering spurious effects, is the same for both, multiplicative and additive mixing. Furthermore, an analysis of the interrelationships for IF-reception leads to comparable conclusions. With respect to IQ-imbalance, the analog front-end can be modelled independently from the utilized technology.

The problem to be solved is how to regenerate the original IQ-signals without knowledge of the front-end transmission, i.e. the mixing matrix \( \mathbf{A} \). For that purpose, adaptive filtering and blind source separation algorithms can be applied (Valkama et al., 2001) as well as several calibration procedures (Hentschel, 2004) for both of the front-end architecture and blind source separation algorithms can be applied i.e. the mixing matrix \( \mathbf{A} \).

3 IQ-Imbalance - effects of amplitude/phase impairments and path mismatches

In the previous sections, we assumed to have only quasi-linear impairments. In reality, the amplitude and phase impairments have a frequency dependent characteristic. Hence, different parts (at different frequencies) of a single channel will be affected differently. Usually, this effect becomes more serious with wider bandwidths. In such cases, the known techniques of IQ-regeneration will have to be translated into deconvolution methods which consider the frequency-selectivity of the imbalance.

Therefore, we need a general description of the wideband signal being affected by amplitude/phase impairments and path mismatches. Following the explanations in (Valkama et al., 2001), we consider a multichannel received signal

\[ s_{RF}(t) = 2 \Re \left\{ z(t) \exp(j \omega_{RF} t) \right\} = z(t) \exp(j \omega_{RF} t) + z^*(t) \exp(-j \omega_{RF} t) \]  

(8)

which is centered at \( \omega_{RF} \) and covers a bandwidth \( B \) (Fig. 2). The operation \((\cdot)^*\) denotes complex conjugation of its argument. In Eq. (8)

\[ z(t) = z_0(t) + z_{IF}(t) \exp(j \omega_{IF} t) + z_{IM}(t) \exp(-j \omega_{IF} t) = z_I(t) + j \cdot z_Q(t) \]  

(9)

is the general multichannel baseband signal.

Without the loss of general validity, the analog components: LO-paths within the frequency down conversion stage, branch filters and amplifiers, analog-to-digital converters (ADC) contribute to the respective frequency-independent or frequency-selective IQ-imbalances. Hence, we can combine the several components what leads to a more general front-end model. In Fig. 3, the amplitude and phase impairments are modelled by the relative errors \( g \) and \( \varphi \), respectively, within the mixing stage. \( H_{NOM}(\omega) \) represents the nominal low-pass-filter (LPF) functionality which rejects the high frequency components after the down conversion. Therefore, the complex signal after the nominal LPF of the branches is:

\[ z(t) = LPF_{NOM} \left\{ s_{RF}(t) : \xi_{LO}(t) \right\} = z_I(t) + j \cdot \left( g \cos(\varphi) z_Q(t) - g \sin(\varphi) z_I(t) \right) = x_I(t) + j \cdot x_Q(t). \]  

(10)

The transfer functions \( H_I(\omega) \) and \( H_Q(\omega) \) stand for the actual mismatch due to the frequency-selective differences of the analog components of the in-phase and the quadrature-phase
branch, respectively. With the use of Eq. (10), we obtain the final IQ-imbalanced signal mixture as:

\[
Z'(\omega) = H_I(\omega) X_I(\omega) + j \cdot H_Q(\omega) X_Q(\omega)
\]
\[
= Z'_I(\omega) + j \cdot Z'_Q(\omega)
\]
\[
= H_I(\omega) Z_I(\omega)
\]
\[
+ j \cdot H_Q(\omega) \cdot \left( g \cos(\varphi) Z_Q(\omega) - g \sin(\varphi) Z_I(\omega) \right)
\]
\[
= \left( H_I(\omega) - j H_Q(\omega) g \sin(\varphi) \right) \cdot Z_I(\omega)
\]
\[
+ j \cdot \left( H_Q(\omega) g \cos(\varphi) \right) Z_Q(\omega)
\]
\[
= A_I(\omega) \cdot Z_I(\omega) + j \cdot A_Q(\omega) \cdot Z_Q(\omega)
\]  

(12)

Thus, Eq. (12) is the generalized version of Eqs. (4) and (7) and can be used for the description of any type of imbalanced IQ-processing front-end.

With the simplifications in (Valkama et al., 2001), the down converted and generally imbalanced signal can be denoted as

\[
Z'(\omega) = G_1(\omega) \cdot Z(\omega) + G_2(\omega) \cdot Z^*(-\omega)
\]  

(13)

with

\[
G_1(\omega) = \frac{1}{2} \cdot \left( A_I(\omega) + A_Q(\omega) \right)
\]
\[
= \frac{1}{2} \cdot \left( H_I(\omega) + H_Q(\omega) g \exp(-j \varphi) \right)
\]  

(14)

and

\[
G_2(\omega) = \frac{1}{2} \cdot \left( A_I(\omega) - A_Q(\omega) \right)
\]
\[
= \frac{1}{2} \cdot \left( H_I(\omega) - H_Q(\omega) g \exp(j \varphi) \right).
\]  

(15)

In Fig. 4, the overlapping is shown of \(Z(\omega)\) and \(Z^*(-\omega)\) due to the IQ-imbalance as given in Eq. (13). This results in the imbalanced multichannel baseband observation \(Z'(\omega)\) shown in Fig. 4b. In general, there are two possible imbalance scenarios: for IF-reception, the image and the desired signal mix onto each other \((S_{IF}^i S_{IM}^i)\) or, within direct conversion, the desired signal is distorted by a complex scaled and complex conjugated version of itself \((S_0)\). Furthermore, it should be mentioned that IQ-imbalance does not equal the distortions of the channel affecting the respective IQ-signals in general. On the one hand, for zero-IF reception \((z(t) = s_0(t))\), the general imbalance description (13) shows comparable transmission behavior like the RF-channel. Hence, in this case, one could include the IQ-regeneration within the channel equalization, at least partly. But, on the other hand, an IF-reception with state-of the art analog IQ-front-ends, can not guarantee a sufficient image-channel-rejection before the first frequency conversion. As result, the image channel will overlap onto the desired channel, which can not be compensated for by channel equalization. An additional IQ-regeneration is required stringently to recreate the orthogonality of the desired and the image channel.

Moreover, we can define a more general version for the signal-to-image-ratio (SIR):

\[
SIR = \frac{|G_1(\omega)|^2}{|G_2(\omega)|^2}.
\]  

(16)

Without a mismatch of the IQ-branches, i.e. \(H_I(\omega) = H_Q(\omega)\), (16) turns into the relation for quasi-linear imbalances which has already been depicted in various articles, e.g. (Valkama et al., 2001; Windisch et al., 2004), etc.

However, in Eq. (13) and Fig. 4, the term \(Z^*(-\omega)\) is caused by the imbalances and represents the image aliasing effects. Therefore, we lose the separability of the respective frequency signals. With practical analog implementations of receiver front-ends, the SIR is usually in the range of \(20 \ldots 40\) dB. For the most of currently utilized receiving methods (zero-IF, low-IF, wideband-IF) and usually processed modulation schemes/orders, this image attenuation is insufficient. In principle, a compensation for IQ-imbalance is needed.

As already mentioned, there are several techniques to obtain the required separation matrix for quasi-linear mixtures. Adaptive filters, blind algorithms as well as test-signal-based calibration methods are state of the art (Valkama et al., 2001; Windisch et al., 2004; Mailand et al., IST 2005).

For the task of retrieving the original signal out of the convolved mixture as depicted in Eq. (13), different deconvolution algorithms were proposed, e.g. (Cardoso et al., 1996). But, since we are aiming at simple, low-power, high-speed and thus cost effective solutions for mobile communications, such deconvolution algorithms turn out not to be suitable candidates to compensate for the frequency-selective

---

Fig. 4. Schematic description of (a) a general multichannel signal and (b) the respective resulting IQ-imbalanced mixture.
IQ-imbalances. The reasons for that are somewhat manifold, e.g.:

- adaptive deconvolution algorithms usually require several hundred up to thousands of symbols to reach their respective equilibrium,

- this convergence process must be went through again after each ’environmental’ change (significant temperature change, reception-frequency changes, etc.), due to the consequent changes within the analog IQ-branches,

- serious mapping of the frequency response of the IQ-imbalances requires digital filters with a sufficient tap-length, whereas the update of each tap has its own underlying adaptive algorithm and

- an additional digital memory will have to be implemented just only to save all coefficients for a large variety of reception situations.

4 Practically required compensation

In order to avoid the expensive deconvolution, we propose to compensate only for the mean values of the general, frequency-selective IQ-imbalance.

Let us consider to have only a relative IQ-branch mismatch, i.e. \( H_1(\omega) = 1 \) and \( H_Q(\omega) = \overline{H}_Q(\omega)/H_1(\omega) \). Hence, we obtain:

\[
G_1 = \frac{1}{2} \cdot \left( 1 + H_Q(\omega) g \exp(-j \varphi) \right) \\
G_2 = \frac{1}{2} \cdot \left( 1 - H_Q(\omega) g \exp(j \varphi) \right)
\]

Further, \( H_Q(\omega) \) is divided into its mean values and its frequency-dependent IQ-deviations for both, amplitude and phase.

\[
H_Q(\omega) = \left( g_Q + D_Q(\omega) \right) \cdot \exp\left( j (\alpha + D_\alpha(\omega)) \right)
\]

with:

\[
g_Q = E\left[ |H_Q(\omega)| \right] \quad \alpha = E\left[ \arg(H_Q(\omega)) \right]
\]

whereas \( E\{\cdot\} \) denotes the expectation or mean value. Standard IQ-regeneration methods compensate for a mixture of \( g \), \( g_Q \), \( \varphi \) and \( \alpha \) for considered quasi-linear amplitude and phase impairments. If one sets the \( D_Q(\omega) = 0 \) and \( D_\alpha(\omega) = 0 \) and thus considers not to have frequency dependencies within the IQ-imbalance, one will obtain a relation for the quasi-linear amplitude and phase impairments within \( G_1 \) and \( G_2 \). Since the respective algebraic expressions do not lead to further insight, we leave it up to the recipient to recalculate the somewhat long formulas if being interested.

Nevertheless, for typical, practical front-end implementations the mean values for phase and amplitude impairments are in the range of \( 1 \ldots 2^\circ \) and \( 1 \ldots 2\% \), respectively. Although there are only a few publications about frequency-dependent amplitude and phase impairments, we can assume to have maximum deviations of \( \pm 0.5^\circ \) and \( \pm 0.03 \) dB per channel within typical GSM front-ends or \( \pm 0.8^\circ \) and \( \pm 0.08 \) dB per channel within IEEE 802.11a (WLAN) front-ends, amongst others deduced from Nakagawa et al. (2004). Besides the sparse citation about such measurements, it seems to be valid to assume values within this range, since almost all commercially available front-ends operate without IQ-imbalance compensation for frequency-selective influences, whereas there are definitely mismatches due to fabrication process variations.

Hence, if we assume to have an arbitrary compensator for the mean values of the IQ-imbalance, the influence of amplitude and phase impairments is comparably marginal and this only for a limited spectral part of the signal. Therefore, we conclude that expensive deconvolutive IQ-regeneration is not a must in order to obtain sufficient signal quality.

5 Simulation results

Computer simulation have been carried out to illustrate the interrelationships of the various IQ-imbalance sources and their compensation with the overall transmission performance of the imbalanced front-end.

Within the simulation, a 16 QAM-modulated signal was transmitted over an additive white gaussian noise (AWGN) channel under the assumption to have an optimal transmitter. The receiver front-end utilizes multiplicative mixing like in Fig. 1a and thus Fig. 3 with a low-IF frequency conversion. The channel bandwidth was set to 20 MHz with a bandgap of 12 MHz, whereas the carrier frequency was \( f_{RF} = 500 \) MHz and the IF was selected at 32 MHz. The transmission was split into blocks of 1024 symbols with each block comprising 64 training symbols for synchronization, etc. The system used a split raised cosine pulse forming with a rolloff factor of 0.35. Within the IQ-paths from the LO, relative impairments for amplitude and phase were included...
with $g = 0.98 (-0.175 \text{ dB})$ and $\varphi = -2^\circ$, respectively. The frequency-selective error in the IQ-branches was modelled by imbalanced LPF.

In Fig. 5, the relative IQ-branch impairments are illustrated for a worst case scenario. As it can be seen, there are additional mean values, i.e. frequency-independent impairments, for the amplitude and the phase with $g_Q \sim 0.99 (-0.087 \text{ dB})$ and $\varphi \sim -1.4^\circ$. The amplitude alternates with variations of $\pm 0.05 \text{ dB}$, which is approximately within the variation range of technical implementations. In order to obtain results at a low-level limit of implementations with almost insufficient matching, we selected the phase to vary with $\pm 1.5^\circ$, what is more than twice the variation within a typical IQ-processing front-end.

Nevertheless, it can be observed that the symbol-error-rate (SER) performance of the whole system reaches acceptable values, if the mean values for all amplitude and phase impairments were compensated for (at $SIR = -20 \text{ dB}$), Fig. 6. Obviously, the overall performance increases as the respective frequency independent impairments are removed. Since the result for the compensation of all mean values of amplitude and phase errors (concerning IQ-imbalance) is near to the ideal, theoretical $SER$-performance for a 16 QAM signal, we can state that no further, expensive deconvolution of the IQ-signals is required. However, if one would do so, the performance would improve further.

The demodulations’ dependence on the initial signal-to-interference ratio ($SIR$) is indicated in Fig. 7. For the given simulation setup, an $SIR$ of $-40 \text{ dB}$ or less frustrates the demodulation almost completely, even for high $SNR$. Starting at $SIR \geq -35 \text{ dB}$, the receiver was able to detect the IQ-signals properly resulting in acceptable $SER$-values.

Based on the presented results, we can conclude that the demodulation of digitally modulated signals reaches a sufficient level for a reasonably badly imbalanced IQ-processing front-end, if the quasi-linear (mean) values of phase and amplitude impairments have been compensated for.

6 Conclusions

Based on the general transmission of direct down-conversion front-ends with respect to in-phase and quadrature-phase signals, a generally valid imbalance model was explained for technology-independent front-ends and arbitrary frequency down conversion. The investigations on frequency-selective components within the IQ-channels led to the result, that the compensation of the mean values of the respective impairments ought to be sufficient for mobile communications receiver.

Acknowledgements. This research was supported by Siemens AG Munich, Communications, Mobile Devices. Additionally the authors would like to thank the staff of the Chair Information Technology for Traffic Systems for inspiring discussions.

References


