Blind IQ-Imbalance Compensation Using Iterative Inversion for Arbitrary Direct Conversion Receivers

Marko Mailand, Raik Richter and Hans-Joachim Jentschel

Dresden University of Technology
Institute of Traffic Information Systems
01069 Dresden, Germany
Email: {mailand, richter, jentschel}@vini.vkw.tu-dresden.de
Telephone: ++49 351 36753 6775

Abstract—Besides low-IF techniques, the direct down conversion or homodyne reception gains more and more attention in academical and industrial research. In this article we describe both, multiplicative and additive mixing within analog direct down conversion front-ends. The effect of gain and phase imbalances of the local oscillator as well as signal path mismatches within the analog front-end lead to a general transmission which can be shown to be the same for multiplicative and additive mixing in general. Based on that, a blind source separation algorithm is described which converges faster and more robust in the presence of additional noise (e.g. due to the channel) than known techniques. The algorithm can compensate the distortions being caused by the impairments within the receiving system without additional knowledge about the front-end or training sequences.

I. INTRODUCTION

Reconfigurability and hence the displacement from analog towards digital signal processing within mobile communication receivers has gained a lot of research interest [1]. The aim is to implement more and more adjustable, digital and not analog components in order to obtain multi-protocol capable systems and to reduce the costs of the receiver, i.e. power consumption, component size or chip area and production costs. In this respect, one concept is to accept spurious effects within the analog domain, what saves the analog effort for compensation, and to eliminate the consequences of these effects digitally. Spurious effects which are addressed in this article are imbalances (phase and amplitude) of the local oscillator (LO) signals and mismatches between different signal paths within an analog receiver front-end. In consequence, there will be imbalances of the respective I- and Q-signals. The RF-signal at the receiver shall be denoted with

\[ s_{RF}(t) = Re\{s(t) \cdot \exp(j \omega_{RF} t)\} \]  

whereas \( s(t) = s_I(t) + j \cdot s_Q(t) \) contains the in-phase (I) and quadrature-phase (Q) component of the modulated signal. The ideal LO-signal can be written as

\[ s_{LO}^{\text{ideal}}(t) = \exp(-j \omega_{LO} t) \cdot \]  

We will investigate the influence of the spurious effects on the base-band signal. This leads to a general relation between conventional direct conversion reception with multiplicative mixing [2] and the alternative concept utilizing additive mixing [3], [4].

The main object of investigation is the regeneration of the transmitted signal which is distorted by the imperfect receiver front-end. Besides test-signal measurements and adaptive filtering, blind source separation (BSS) algorithms can be utilized to regenerate the wanted signals [5]. Signal regeneration with the use of BSS algorithms avoids additional, expensive calibration procedures with test-signal measurements.

Starting with a general description of the direct conversion front-end transmission, the use of an appropriate blind source separation algorithm is presented [6] and compared to other BSS concepts. The application of a fast and robust iterative inversion BSS approach will be shown.

II. DIRECT DOWN CONVERSION FRONT-ENDS

The technique of direct down conversion (DDC) transforms the RF-signal directly down to base-band. 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 \( \psi_i \) and amplitude \( g_i \) imbalances. Therefore, the ideal LO-signal changes into

\[ s_{LO}(t) = g_1 \cdot \cos(\omega_{LO} t + \psi_1) - j \cdot g_2 \cdot \sin(\omega_{LO} t + \psi_2) \]  

Fig. 1. DDC front-end utilizing multiplicative mixing; with imbalance of the LO-signal and path mismatch; the grayly grounded elements only model the error sources - they are not discrete components.
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.

A. Direct Conversion by 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 $s_{BB}(t) = LPF\{s_{RF}(t) \cdot s_{LO}(t)\}$. The first signal path is assumed not to be matched to the second path. This leads to a relative path mismatch $M_i$. In Fig. 1, 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. 1 and considering all the impairments, we get:

$$x_1(k) = \frac{1}{2} g_1 M_1 \left( s_I(k) \cos(\psi_1) + s_Q(k) \sin(\psi_1) \right)$$ (4)

$$x_2(k) = \frac{1}{2} g_2 \left( -s_I(k) \sin(\psi_2) + s_Q(k) \cos(\psi_2) \right).$$ (5)

In general, the equations (4) and (5) can be rewritten in a matrix form to consider all possible, linear imbalances (except DC-offsets):

$$x = A \cdot s = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \end{bmatrix} \begin{bmatrix} s_I(k) \\ s_Q(k) \end{bmatrix}. $$ (6)

We consider to have arbitrary gain and phase impairments within each of the IQ-paths, although a relative IQ-imbalance was sufficient [5]. Hereupon, we define the signal-to-interference ratio (SIR) for the I-path and Q-path:

$$\text{SIR}_I = 20 \cdot \log_{10} \left| \frac{A_1}{B_1} \right| \quad \text{and} \quad \text{SIR}_Q = 20 \cdot \log_{10} \left| \frac{B_2}{A_2} \right|. $$ (7)

B. Direct Conversion by 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 communication receiver the former six-port was reduced to a five-port structure. Fig. 2 shows the general architecture for additive DDC utilizing a five-port. Clearly, the mixers (multiplicative elements in Fig. 1) are replaced by appropriate summing elements. 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 [3].

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

$$x_i^{LPF}(t) = G_i^2 M_i + M_i \cdot (s_I^2(t) + s_Q^2(t))$$ (8)

rectified wave including DC-offsets

$$\downarrow 2 M_i \cdot (s_I(t) G_i \cos(\varphi_i) + s_Q(t) G_i \sin(\varphi_i))$$

wanted signal components

The phase shift elements $\varphi_i$ 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 [7] or by the implementation of an appropriate front-end architecture [4]. Therefore, we obtain measurements which consist only of linear mixtures of the IQ-signals, i.e. $x_i(k) = 2 M_i \cdot (s_I(k) G_i \cos(\varphi_i) + s_Q(k) G_i \sin(\varphi_i))$ in (8). In consequence, the transmission of the DDC front-end with additive mixing is determined by

$$x = A \cdot s = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \end{bmatrix} \begin{bmatrix} s_I(k) \\ s_Q(k) \end{bmatrix}. $$ (9)

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 $A$.

The composition of the base-band 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 $x_i(k)$ 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.
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 $A$. For that purpose, blind source separation algorithms can be applied [5], [7] for both of the front-end architectures.

III. BLIND SOURCE SEPARATION (BSS)

Blind source separation has already a lot of applications in field of communication e.g., IQ-imbalance correction within IF-receivers [5], multiuser detection in MIMO systems [8], etc. There are already suggestions to apply BSS to blind front-end parameter estimation within direct conversion e.g., for the special case of additive mixing by the six-port-based technology in [4], [7].

The problem of blind separation of sources consists in retrieving unknown sources $s$ from only observing a mixture $x$ of them, i.e.:

$$x = As.$$  \hspace{1cm} (10)

Clearly, the fundamental BSS problem (10) corresponds to the direct down conversion transmission which considers spurious distortions due to the LO-imbalances and the path mismatches. Hence, there are no prerequisites for the relative power level of the LO-imbalances and the path mismatches.

The separating matrix $B$ is statistically independent of one another. The separating matrix $B$ is obtained in terms of Kullback-Leibler divergence.

$$B = \text{opt}$$

The separating matrix $B$ is obtained in terms of Kullback-Leibler divergence. The separating matrix $B$ is obtained in terms of Kullback-Leibler divergence.

$$B = \text{opt}$$

A. Iterative Inversion

The iterative inversion (II) algorithms were derived by exploiting special estimates: robust estimates [6]. II algorithms can be interpreted as iterative correction of the current estimate $B^{(k)}$ of the mixing system $B^{-1}$ in the direction of the robust estimate $A(B)$ [6]. By definition we have a cost function $\psi(y)$ which is obtained in terms of Kullback-Leibler divergence. Furthermore, it can be shown that a necessary condition for the optimal separation matrix $B_{\text{opt}}$ is obtained if the derivation of the cost function with respect to $B$ is [6]:

$$\frac{\partial \psi(y)}{\partial B} \bigg|_{B=B_{\text{opt}}} = (R_{\psi x} - B^{-T})|_{B=B_{\text{opt}}} = 0$$  \hspace{1cm} (14)

$R_{\psi x} = E\{\psi(y)x\}$ is the correlation matrix between a nonlinear function of the outputs $\psi(y)$ and the input $x$. $\psi(\bullet)$ describes the nature of the mixing system $A$. The optimum nonlinear function $\psi(\bullet)$ is composed of component-wise derivations of probability density functions of the separation outputs [6]. However, the exact nonlinear function $\psi(\bullet)$ is usually unknown and has to be estimated.

Generally speaking, not the direct gradient of the linear correlation implies the fastest way to the separation optimum (like with the well-known LMS algorithm) but the gradient within the respective nonlinear space of the mixing problem. By including a nonlinearity which corresponds sufficiently to the respective mixing problem, the convergence behavior as well as the convergence speed can be improved. So, for communication signals, cubic nonlinearities are preferred since they allow for computational efficiency and especially for the fact of digitally modulated signals (PSK, QAM) having negative kurtosis [5].

On the one hand, this leads to a BSS algorithm with comparably good performance for robust estimates. On the other hand, a BSS algorithm with less computational effort and faster convergence can be derived. Based on that, Cruces-Alvarez et al. [6] developed an iterative inversion BSS algorithm for online-learning with the update rule

$$H(y^{(k)}) = f(y)g^T(y) - I$$  \hspace{1cm} (15)

and the normalized adaptive step-size

$$\mu^{(k)} = \frac{\eta}{1 + \eta |g^T(y)f(y)|}.$$  \hspace{1cm} (16)

The resulting normalized II algorithm for the separation matrix is given by [6]:

$$B^{(k+1)} = \left( I - \frac{f(y)g^T(y) - I}{1 + \eta |g^T(y)f(y)|} \right) B^{(k)}$$  \hspace{1cm} (17)

with $I$ as identity matrix and $\eta \in (0, 1)$ as step-index. In general, $f(y)$ and $g(y)$ are arbitrary nonlinear functions which act componentwise on their arguments. The advantage of using two nonlinearities instead of one [5], [9] is the robustness to the noise present in the mixture.

For the sake of computational efficiency and further including the expected property of the source signals to have negative kurtosis, we define the algorithm core functions as:

$$f(y_i) = |y_i|^2 y_i$$  \hspace{1cm} (18)

$$g(y_i) = \text{sign}(\text{Re}\{y_i\}) + j \cdot \text{sign}(\text{Im}\{y_i\})$$  \hspace{1cm} (19)

Further, the II algorithm of (17) performs independent of the source signals relative powers due to its equivariance property. Hence, there are no prerequisites for the relative power level distortions due to the LO-imbalances and the path mismatches.
B. Convergence Performance of BSS

In many applications of blind source separation a measure for the respective algorithms’ convergence is required. Besides signal-to-interferer ratios (SIR) like in [5], one can investigate the overall mean covariance for evaluating the convergence performance. That is to calculate the covariances $E\{\Delta \mathbf{C}^{(k)}\}$ of the entries of the relative error matrix after the algorithm has converged, where

$$\Delta \mathbf{C}^{(k)} = \Delta \mathbf{B}^{(k)} \Lambda = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix} \quad (20)$$

with $\Delta \mathbf{B}^{(k)} = \mathbf{B}^{(k)} - \mathbf{B}_{opt}$. (21)

$\Delta \mathbf{C}^{(k)}$ is convenient for our analysis because the recovering error is written as $\Delta s = \Delta \mathbf{B}^{(k)} \mathbf{x} = \Delta s$. Then, the overall mean covariance is the geometrical mean:

$$\sigma_{err}^2 = \sqrt{(\sigma_{11}^2 + \sigma_{21}^2)^2 + (\sigma_{12}^2 + \sigma_{22}^2)^2} \quad (22)$$

whereas $\sigma_{ij}^2 = E\{(\epsilon_{ij} - E[\epsilon_{ij}])^2\}$.

(23)

The measure $\sigma_{err}^2$ only enables the comparison of the remaining deviation of the respective BSS algorithm, i.e. the precision of convergence. The quality of separation has to be measured differently e.g., by using SIR or error rate performance.

IV. APPLICATION OF BSS TO DIRECT DOWN CONVERSION FRONT-END PARAMETER ESTIMATION

For detailed investigations on the performance of a direct receiver that utilizes the II algorithm (17), a communication system was analyzed. Within the system, a 16QAM modulated signal was transmitted over an AWGN-channel. The pulse shaping raised cosine filter has a rolloff of 0.3. The LPF of the front-end are implemented as 16th order FIR filter. The carrier frequency is 1GHz and we use transmission blocks of 1024 bits with prepended training sequences of 256 bits. As front-end, we use a five-port with additive mixing, Fig. 2 with: $G_1 = 0.7$, $G_2 = 1$, $G_3 = 0.8$, $\varphi_1 = 10^\circ$, $\varphi_2 = 0^\circ$, $\varphi_3 = 22^\circ$, $M_1 = 0.92$, $M_2 = 1$ and $M_3 = 0.97$, which leads to $\text{SIR}_1 \approx 15\text{dB}$ and $\text{SIR}_Q \approx 7.7\text{dB}$. The simulation results of this parameter setup are approximately the same, if the multiplicative mixing front-end was chosen with $g_1 = 0.7$, $g_2 = 0.8$, $\psi_1 = 10^\circ$, $\psi_2 = -65^\circ$ and $M_1 = 0.95$. Hence, the relative amplitude and phase errors for the multiplicative mixing front-end were 10% and 5°, respectively.

Because the system transmissions of both of the direct conversion techniques are equal, it is sufficient to analyze the BSS algorithm with respect to one of the DDC front-ends.

The results of the BSS algorithm are used to regenerate the IQ-constellation which is distorted by the front-end. The step-index parameter of the II algorithm is set to $\eta = 2.782 \cdot 10^{-3}$. Further, it is required to attach a phase synchronization after the BSS; we utilize a basic trained phase synchronization technique.

If we compared the convergence measure of II to the EASI BSS algorithm (step-index: $\lambda = 2.7 \cdot 10^{-3}$) which was suggested for IQ-imbalance compensation in [5], we observe the same convergence precision of the two algorithms, Fig. 3. Therefore, the algorithms themselves add the same amount of indeterminacy, i.e. variation to the IQ-signals. Hence, we can compare the quality of separation by means of symbol-error-rate (SER) to evaluate which algorithm is more suitable to IQ-imbalance compensation.

In Fig. 4, the ‘theoretical boundary’ is the optimum SER performance for a 16QAM modulated signal. It corresponds to the direct calculation under the assumption that the front-end parameters were known explicitly. Thus, the IQ-regeneration with respect to source separation is perfect in this case. As it can be seen, the II algorithm performs slightly better than the EASI algorithm in the presence of stronger noise. This is due to the utilization of two nonlinearities, which make the II more robust. Since the precision of convergence is the same for both algorithms, the II generates better estimates for the demixing matrix $\mathbf{B}$.

Nevertheless, neither the II nor the EASI algorithm reach the optimum of separation, since the remaining deviations at the steady state. One could improve the separation performance by decreasing the step-index parameters. Furthermore, the step-index parameters can be selected such that the EASI reaches the performance of the II. By doing so, the convergence time of the EASI will significantly increase and $\sigma_{err}^2$ will decrease. Besides, the computational complexity of the II is less, since...
The value for \( k \) of its steady state mean (the EASI algorithm within the given setup is the coefficients of \( C \) of front-end cooling scheme speed of a BSS algorithm can be additionally accelerated by sources without a circular relocation (by means of phase-fore, it can only perform an orthogonalization of respective of the initial vector constellation of the source signals. There-

reason is that in general a BSS algorithm has no knowledge converge to the identity matrix that the steady state mean of the global system does not faster than the EASI algorithm. Moreover, one can observe SNR=30dB. algorithm (as reference: the time point of convergence II algorithm has for the EASI face one sign operation within the II algorithm has\( k_c,\text{EASI} \approx 1000 \). The value for \( k_c \) is defined as the time point when the last coefficient of the global system \( C^{(k)} \) enters the \( \sigma^2_{ij} \)-region of its steady state mean \( k_c = \max\{k_c,\text{i}\} \forall i \). In Fig. 5 the coefficients of \( C^{(k)} \) are shown. As it can be seen, the II algorithm has \( k_{c,\text{II}} = 580 \). Hence, it converges much faster than the EASI algorithm. Moreover, one can observe that the steady state mean of the global system does not converge to the identity matrix I, what one could expect. The reason is that in general a BSS algorithm has no knowledge of the initial vector constellation of the source signals. Therefore, it can only perform an orthogonalization of respective sources without a circular relocation (by means of phase-synchronization for communication signals). The convergence speed of a BSS algorithm can be additionally accelerated by a cooling scheme for the respective step-index parameter \([9]\). That is an appropriate step-index sequence which decreases by time: \( \{\eta, \lambda\} \propto 1/k \).

In Fig. 6, the capability of the II algorithm is shown to adapt to a changing environment. It is assumed that the front-end transmission and hence the mixing matrix \( A \) change abruptly since e.g., the receiver performs a frequency hopping and therefore the imbalances change. In the example in Fig. 6 at \( k = 1200 \), we changed the gain and phase parameters for the DDC front-end as follows: \( G_1 = 0.7 → 0.8, G_2 = 0.8 → 0.9, \varphi_1 = 10^\circ → 10.3^\circ \) and \( \varphi_2 = 22^\circ → 21.5^\circ \). Obviously, the II algorithm adapted to the new mixing matrix within a new convergence time \( k' \approx 100 \). Because of only a comparably small change from the steady state of the separation matrix \( B \), the new \( k' \) is significantly less than the initial \( k_c \).

In comparison to the EASI, the II algorithm converges faster, more precise in the presence of strong noise and has less computational complexity.

V. CONCLUSION

The general relation between direct down conversion techniques with multiplicative and additive mixing was carried out. With the consideration of imbalances and mismatches of a real front-end, it was shown that the general signal transmission is the same for both techniques. The resulting front-end transmission corresponds to the basic problem formulation of any blind source separation (BSS). Therefore, BSS algorithms can be utilized to cancel the effects of the erroneous front-end.

With this respect, a fast and robust BSS algorithm (II) was described and used for IQ-imbalance compensation. The influences of noise and non-stationary environments were investigated and compared to the EASI algorithm.

ACKNOWLEDGMENT

This research was supported by Siemens AG Munich, Communications, Mobile Devices. Additionally the authors would like to thank the staff of the Communication Systems Design Group for inspiring discussions.

REFERENCES


