Transceiver inphase/quadrature-imbalance, ellipse fitting, and the universal software radio peripheral

Cristian R. Rojas, Per Zetterberg, and Peter Händel, Senior Member, IEEE

Abstract—In this paper we introduce a method for IQ imbalance parameter estimation based on ellipse fitting. The performance of the method is analytically derived. In particular, it is shown that the method exhibits a small bias (which can be negligible under some standard practical conditions) and a variance slightly above the Cramér-Rao bound. The method is then applied to measurements from a contemporary BiCMOS transceiver which is used on one of the most popular daughterboards of the universal software peripheral (USRP). In our measurements the phase skew varies up to five degrees with the base-band frequency, while the amplitude imbalance varies between 0 – 0.3 dB over carrier frequencies and across hardware units. The time variation however is only 0.004 dB in amplitude and 0.06 degrees in phase. This indicates that the units could either be calibrated on-line when there is no transmission (in a two antenna MIMO system one antenna could transmit a calibration signal to the other), or they could be calibrated during production, in which case a table with different carrier and base-band frequencies would be needed. However, there is no need to estimate the parameters on every burst.

Index Terms—inphase/quadrature (IQ) imbalance, Cramér-Rao bounds, ellipse fitting, universal software radio peripheral USRP.

I. INTRODUCTION

TIME-EFFICIENT test methods are a requirement for effective large-volume production tests of electronic devices. Radio transceivers are a pervasive component in consumer and industrial electronics produced in immense volumes, for example mobile phones, wireless local area network routers, remote controls, and so on.

This work considers a fully digital and software implementable time-efficient method for determination of the inphase/quadrature (IQ) imbalance of contemporary direct-conversion radio frequency transceivers based only on the digital baseband output. The method is derived from the practice of measuring the imbalance by manual reading of elliptic Lissajous plots presented by a dual beam cathode ray tube (CRT) oscilloscope [1]. Considering a receiver, a perfectly balanced one should display a circle centered on the CRT-screen, whereas practical receivers display an ellipse because of the gain imbalance between the I and Q channels, and the quadrature skew or phase offset. The center of the ellipse is offset from the center of the display due to the leakage from the local oscillator (LO).

The purpose of this paper is threefold. First, the performance of the popular method in [2] is analyzed in detail in an attempt to characterize the earlier observed systematic error, as well as the accuracy of the same method. In addition, the performance of the method is compared with objective bounds on the achievable performance [16], quantifying the loss in performance compared with statistically efficient methods – the price paid for the low numerical complexity of the employed method.

Secondly, the ellipse fitting approach is used to characterize the IQ imbalance of a contemporary BiCMOS radio frequency transceiver, namely, the MAXIM 2829 (www.maxim-ic.com/datasheet/index.mvp/id/4532/t/al). The setup is designed to

Fitting the parameters of an ellipse to noisy sampled data is a problem of great interest in several communities, including two dimensional image processing [2], medical imaging [3], and computer vision [4], [5], [6], [7], [8]. A seminal contribution to the algorithmic development of robust ellipse fitted was Fitzgibbon’s method [2]. The ingenuity of the method in [2] is the relaxation of the elliptic constraint imposed on the solution, leading to a problem with a closed form solution. From an instrumentation and measurement point of view, the method in [2] is a way to replace a manual reading of an elliptic Lissajous plot [9], by automatic digital signal processing of A/D converter samples. Within the instrumentation and measurement community, the fitting of an ellipse makes sense, as e.g. the work on estimation of particle size and velocity in laser anemometry [10], [11], and impedance determination [12], [13], [14], [15]. Some numerical characterizations of the ellipse fitting in [2] to the problem of impedance characterization were presented in [12], [13] highlighting some systematic or bias errors, especially for scenarios with low signal-to-noise ratios. One of the contributions of this work is a detailed analysis of this systematic error.

For scenarios characterized by an additive noise model, a detailed theoretic performance assessment was performed in [16]. In [16], under a Gaussian assumption, the ultimate performance of any unbiased estimator was addressed by the Cramér-Rao lower bound [17]. In addition, the performance of nonlinear least squares estimation of the sought parameters (for example, the method in [18] which is a generalization of the IEEE standardized tone fit algorithm to dual channels, [19]) was investigated. Nonlinear least-squares estimation to characterize IQ-imbalance of direct-conversion receivers was considered in [20], revealing not only the excellent performance of the nonlinear least-squares method in general, but also its outlier performance under stressed situations aiming at reducing the testing time.

For scenarios characterized by an additive noise model, a detailed theoretic performance assessment was performed in [16]. In [16], under a Gaussian assumption, the ultimate performance of any unbiased estimator was addressed by the Cramér-Rao lower bound [17]. In addition, the performance of nonlinear least squares estimation of the sought parameters (for example, the method in [18] which is a generalization of the IEEE standardized tone fit algorithm to dual channels, [19]) was investigated. Nonlinear least-squares estimation to characterize IQ-imbalance of direct-conversion receivers was considered in [20], revealing not only the excellent performance of the nonlinear least-squares method in general, but also its outlier performance under stressed situations aiming at reducing the testing time.

The purpose of this paper is threefold. First, the performance of the popular method in [2] is analyzed in detail in an attempt to characterize the earlier observed systematic error, as well as the accuracy of the same method. In addition, the performance of the method is compared with objective bounds on the achievable performance [16], quantifying the loss in performance compared with statistically efficient methods – the price paid for the low numerical complexity of the employed method.

Secondly, the ellipse fitting approach is used to characterize the IQ imbalance of a contemporary BiCMOS radio frequency transceiver, namely, the MAXIM 2829 (www.maxim-ic.com/datasheet/index.mvp/id/4532/t/al). The setup is designed to
allow the characterization of the transmitter as well as the receiver. We investigate the IQ imbalance as a function of base-band frequency, carrier frequency, time and across five different hardware units. We find that the IQ imbalance is clearly varying with all the above parameters except time. This indicates that the units could be calibrated during production in which case a table with different carrier frequencies would be needed. Another alternative would be to do calibration when the unit is idle.

Contemporary and future wireless communication systems put high demands on accurate and time-efficient test methods for production and product validation, especially for multiple input multiple output (MIMO) systems, where each branch is equipped with its own transceiver. A suitable method should provide estimates with low bias and variance while at the same time require a limited amount of computations. As indicated in [12], ellipse fitting is computationally less demanding than brute-force nonlinear least-squares methods.

Thirdly, the MAXIM 2829 is used in one of the most popular daughterboards of the universal software radio peripheral (USRP). The USRP has entered the research community as a versatile radio platform for education, development and research [21], with applications aiming at navigation [22], cognitive radio [23], and software defined radio [24]. Thus the results obtained herein have an interest of their own, owing to the widespread use of the USRP.

The paper is organized as follows. In Sec. II, the problem is formulated, and the state-of-the-art is reviewed. The performance analysis is performed in Sec. IV. The evaluation on a set of MAX2829 is presented in VI. Finally, the conclusions are drawn in Sec. VII.

II. PROBLEM SET-UP AND STATE-OF-THE-ART

In the subsequent paragraphs, the measurements are introduced, different parameterizations of the signal model are discussed, and benchmarks bounds and methods are shortly reviewed.

A. Measurements

Consider the dual-channel measurements

\[ y_1(n), \ldots, y_1(N), y_2(n), \ldots, y_2(N). \]  

(1)

where the subscript denotes the channel number, and \( n \) denotes the running time index, \( n = 1, \ldots, N \). Thus, in total there are \( 2N \) recorded samples, which form the basis for extracting the parameters of interest. These parameters depend on the application in mind, as outlined in the sequel.

B. A Parametric Model

A parametric model of the measurements is considered, where the unknown parameters are gathered in the parameter vector \( \theta \), that is

\[ y_1(n; \theta) = s_1(n; \theta) + v_1(n) \]  

(2)

and

\[ y_2(n; \theta) = s_2(n; \theta) + v_2(n). \]  

(3)

In (2)-(3), \( s_k(n; \theta) \) (for \( k = 1, 2 \)) denotes the undistorted output, and \( v_k(n) \) a noise term, including additive thermal noise, model imperfections, quantization, harmonic distortion, and so on. The source signal in the first channel is modeled by

\[ s_1(n; \theta) = A_1 \cos(\omega_0 n) + B_1 \sin(\omega_0 n) + C_1 \]  

(4)

where \( A_1 \) and \( B_1 \) are the amplitudes of the in-phase and quadrature components and \( C_1 \) is the DC-level. The quantity \( \omega_0 \) denotes the normalized angular (base-band) frequency, that is \( \omega_0 = 2\pi F/F_S \), where \( F \) is the absolute frequency in Hertz and \( F_S \) the sampling frequency. In a similar vein, the parametric model for the second channel reads

\[ s_2(n; \theta) = A_2 \cos(\omega_0 n) + B_2 \sin(\omega_0 n) + C_2. \]  

(5)

As noted in (4) and (5), the sinusoidal frequency \( \omega_0 \) is common to both channels. Hence, the parameter vector \( \theta \) with the seven unknown parameters is defined as

\[ \theta := [A_1 \ B_1 \ C_1 \ \omega_0 \ C_2 \ B_2 \ A_2]^T \]  

where \( T \) denotes the transpose operation.

C. Alternative parameterizations and their relevance

In many applications (as outlined later on), the interest lies in the estimations of the \( A_k, B_k \) and \( C_k \) and transformations thereof, and thus \( \omega_0 \) may be considered as a nuisance parameter. An alternative parameterization includes the amplitude-phase model

\[ y_1(n; \theta) = \alpha_1 \cos(\omega_0 n) + C_1 + v_1(n) \]  

(6)

\[ y_2(n; \theta) = \alpha_2 \cos(\omega_0 n + \phi_\Delta) + C_2 + v_2(n) \]  

(7)

with \( \alpha_k = \sqrt{A_k^2 + B_k^2} \). In (6), we have fixed the initial phase of the first channel, and introduced the phase difference \( \phi_\Delta \) in (7). For later reference, let us the parameter vector

\[ \vartheta := [\alpha_1 \ \alpha_2 \ \phi_\Delta \ C_1 \ C_2]^T. \]  

(8)

Other parameterizations include relative amplitude differences \( \alpha_\Delta := \alpha_1 - \alpha_2 \) and quotients \( \alpha_{11} := \alpha_2/\alpha_1 \). In particular: i) for phase doppler anemometry the parameters of interest are the angular frequency \( \omega_0 \), which is proportional to the velocity, and the phase difference \( \phi_\Delta \) that determines the particle size [10], [11], ii) for impedance measurements the parameters of interest are the phase difference \( \phi_\Delta \) and amplitude quotient \( \alpha_{11} \) [12], [13], and iii) for mixer imbalance measurements [20], [25], the parameters of interest are:

- The gain imbalance
  \[ G := \alpha_{11} = \frac{\alpha_2}{\alpha_1}. \]  

(9)

- The quadrature skew
  \[ Q := \frac{\pi}{2} - \phi_\Delta. \]  

- The LO leakage
  \[ L := 2\frac{C_1^2 + C_2^2}{\alpha_1^2 + \alpha_2^2}. \]  

(10)

One may note that the model in (6)-(7) is slightly less general than (2)-(5) because of the alignment of the initial phases. For the problem at hand, however, it imposes no restrictions.
D. Benchmarks: ML, NLS and CRB

Let the error terms in (2)–(3) be modeled as jointly independent zero mean Gaussian white noises with variances $\sigma_1^2$ and $\sigma_2^2$, respectively. In the Gaussian scenario, the method of maximum likelihood (ML) is given by the minimizer $\hat{\theta}_{\text{ML}}$ of the least-squares criterion, which is [11]

$$\hat{\theta}_{\text{ML}} = \arg \min_{\theta} \sum_{k=1}^{2} \frac{1}{N} \sum_{n=1}^{N} \left(y_k(n) - s_k(n; \theta)\right)^2.$$

Following the methodology in [11], an estimator may be derived. The performance in terms of error variance of such an estimator is expected (at least as $N \to \infty$) to coincide with the Cramér-Rao bound (CRB) on error performance [17]. However, the ML estimator depends on the unknown noise variances. In [16], the nonlinear least-squares estimator $\hat{\theta}_{\text{NLS}}$ was investigated:

$$\hat{\theta}_{\text{NLS}} = \arg \min_{\theta} \sum_{k=1}^{2} \frac{1}{N} \sum_{n=1}^{N} \left(y_k(n) - s_k(n; \theta)\right)^2.$$

The study in [16] revealed that the estimates were obtained with an accuracy that basically coincides with the CRB for a wide span of channel signal-to-noise ratios. It is obvious that the numerical complexity obtaining the latter estimate is lower than for the ML, because of the independence of the noise variances. In summary, the nonlinear-least-squares estimate $\hat{\theta}_{\text{NLS}}$ is favorable over $\hat{\theta}_{\text{ML}}$. One may note that $\hat{\theta}_{\text{NLS}}$ is studied in some detail in [18], where in particular an efficient algorithm for its implementation is proposed. One may note the similarities between the seven-parameter fit algorithm presented in [18] and the IEEE standardized four-parameter fit for the single channel case [19]. The excellent performance of the algorithm in [18] is also demonstrated in [16].

The CRB is an objective bound on the achievable estimation error variance of any method [17]. In practice, it is feasible to derive the bound for unbiased estimators under the Gaussian assumption, as shown in [19]. This result was further refined in [26] following the derivations in the classical paper [27]. The generalization of the CRB to the seven-parameter model (2)-(5) was derived in [16]. Transformed values for $G$, $Q$ and $L$ according to (9)–(10) are found in [20]. For easy reference, the key results on the CRB are presented in Table I. Here, the signal-to-noise ratio (SNR) per measurement channel is defined as (for $k = 1, 2$)

$$\text{SNR}_k := \frac{A_k^2 + B_k^2}{2\sigma_k^2} = \frac{\alpha_k^2}{2\sigma_k^2}.$$ (11)

In our work, the CRB results in Table I serve as baseline for the employed ellipse fitting algorithm for the problem at hand. It is worth remarking that similar CRB results have been derived in [28], [29], [30] for curve and surface estimation, where the time information associated with the measurements is not available (i.e., the frequency $\omega_0$ becomes a nuisance parameter).

III. ELLIPSE FITTING

In this section we revisit the ellipse fitting algorithm of [2]. To this end, we first introduce some definitions and assumptions. Consider two sinusoidal signals given by (6)-(7), where $\alpha_1 \in \mathbb{R}^5$, $\alpha_2 \in \mathbb{R}^5$, $\phi_\Delta \in (-\pi, \pi) \setminus \{0\}$, $C_1 \in \mathbb{R}$ and $C_2 \in \mathbb{R}$ are unknown constants to be estimated, and $\omega_0 \in \mathbb{R}^+$ is an unknown nuisance parameter. The sequences $\{v_1(n)\}_{n \in \mathbb{N}}$ and $\{v_2(n)\}_{n \in \mathbb{N}}$ are jointly independent Gaussian white noise signals of zero mean and variances $\sigma_1^2$ and $\sigma_2^2$, respectively.

The ellipse fitting algorithm is based on the observation that the noiseless points $\{(y_1(n) - v_1(n), y_2(n) - v_2(n))\}_{n \in \mathbb{N}}$ belong to an ellipse [2].

A. Ellipse parameters

From (6)-(7) we have that

$$\left(\frac{y_1 - v_1 - C_1}{\alpha_1}\right)^2 + \left(\frac{[y_1 - v_1 - C_1] \cos(\phi_\Delta)}{\alpha_1 \sin(\phi_\Delta)} - \frac{y_2 - v_2 - C_2}{\alpha_2 \sin(\phi_\Delta)}\right)^2 = 1$$

where we have dropped the argument $n$ for simplicity. This can be further simplified to yield a quadratic form in $x := y_1 - v_1$ and $y := y_2 - v_2$ of the form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$ (12)

where

$$a := \alpha_2^2$$
$$b := -2\alpha_1 \alpha_2 \cos(\phi_\Delta)$$
$$c := \alpha_1^2$$
$$d := 2\alpha_1 \alpha_2 \cos(\phi_\Delta) C_2 - 2\alpha_2^2 C_1$$
$$e := 2\alpha_1 \alpha_2 \cos(\phi_\Delta) C_1 - 2\alpha_1^2 C_2$$
$$f := \alpha_2^2 C_1^2 + \alpha_1^2 C_2^2 - 2\alpha_1 \alpha_2 \cos(\phi_\Delta) C_1 C_2 - \alpha_1^2 \alpha_2^2 \sin^2(\phi_\Delta).$$

This set of equations defines a mapping from the parameters in $\theta$ of (6)-(7) to the coefficients

$$\eta := [a \ b \ c \ d \ e \ f]^T$$ (13)
of the quadratic form (12). Notice that, since (12) defines an ellipse, it holds that \( f \neq 0 \) and \( f(b^2 - 4ac) > 0 \).

The mapping (12) is not invertible, since (12) is invariant with respect to multiplication by a nonzero constant. To solve this issue, we can scale (12) so that \( 4ac - b^2 = 1 \) (which imposes \( f < 0 \)). Under such condition, we arrive at the inverse mapping:

\[
\begin{align*}
\alpha_1 &= \sqrt{kc} \\
\alpha_2 &= \sqrt{ka} \\
\cos(\phi_\Delta) &= -\frac{b}{2\sqrt{ac}}
\end{align*}
\]

\( C_1 = be - 2cd \)

\( C_2 = bd - 2ac; \quad k = 4(cd^2 - bde + ac^2 - f). \)

The relations between the six parameters gathered in \( \eta \) in (13) and five ones gathered in \( \vartheta \) in (8) are given by

\[
\begin{align*}
a &= \frac{\alpha_2}{2|\alpha_1| \sin(\phi_\Delta)} \\
b &= -\frac{\cos(\phi_\Delta)}{|\sin(\phi_\Delta)|} \\
c &= \frac{\cos(\phi_\Delta)}{\alpha_1} \\
d &= \frac{\cos(\phi_\Delta)}{|\sin(\phi_\Delta)|} C_2 - \frac{\alpha_2}{\alpha_1 |\sin(\phi_\Delta)|} C_1 \\
e &= \frac{\cos(\phi_\Delta)}{|\sin(\phi_\Delta)|} C_1 - \frac{\alpha_1}{\alpha_2 |\sin(\phi_\Delta)|} C_2 \\
f &= \frac{\alpha_2}{2|\alpha_1| \sin(\phi_\Delta)} C_1^2 + \frac{2\alpha_2}{|\sin(\phi_\Delta)|} C_2^2 - \frac{\cos(\phi_\Delta)}{|\sin(\phi_\Delta)|} C_1 C_2 - \frac{\alpha_1 \alpha_2}{2} |\sin(\phi_\Delta)|. 
\end{align*}
\]

B. Algorithm

We now have the problem of how to fit an ellipse to the data (1). For easy reference, the method in [2] is shortly reviewed below. Form the matrices

\[
\begin{align*}
D &:= \begin{bmatrix}
y_1^1(1) & y_1^2(1) & y_2^2(1) & y_1(1) & y_2(1) & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
y_N^1(N) & y_N^2(N) & y_2^2(N) & y(1) & y_2(1) & 1
\end{bmatrix} \\
C &:= \begin{bmatrix}
0 & 0 & 2 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{align*}
\]

and the parameter vector \( \eta \) in (13). Then, estimate \( \eta \) as

\[
\hat{\eta} = \arg \min_\eta \eta^T D^T D \eta \quad \text{s.t.} \quad \eta^T C \eta = 1. \quad (17)
\]

The solution of (17) can be obtained by computing the generalized eigenvectors of \( (D^T D, C) \), i.e., by solving the equation \( D^T D x = \lambda C x \) for \( \lambda \) and \( x \), subject to the constraint \( x^T C x = 1 \). As it was shown in [2], there is a unique positive solution for \( \lambda \), and \( \hat{\eta} \) corresponds to the solution \( x \) associated with such \( \lambda > 0 \). Finally, the estimate \( \hat{\vartheta} \) (c.f. (8)) can be computed from \( \hat{\eta} \) using the relations (14).

IV. PERFORMANCE ANALYSIS

We are first interested in the accuracy of \( \hat{\vartheta} \) in (8) when ellipse fitting is applied. The performance of the other parameters (c.f. Table I) are then obtained by proper transformations.

In order to study the performance of the ellipse fitting algorithm, we proceed in three steps: i) Compute the asymptotic behavior of the matrix \( D^T D \) in (17), ii) determine the asymptotic behavior of the estimator \( \hat{\eta} \) in (17), and iii) calculate the asymptotic behavior of the parameters \( \vartheta \) from \( \hat{\eta} \).

In order to simplify the analysis, we can use the fact, already noted in [2], that the ellipse fitting algorithm is invariant with respect to translations. In fact, if \( y_1 \) and \( y_2 \) are replaced by \( y_1 + d_1 \) and \( y_2 + d_2 \) for some constants \( d_1 \) and \( d_2 \), respectively, then the effect is equivalent to post-multiply \( D \) by a matrix

\[
H := \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
2d_1 & d_2 & 0 & 1 & 0 & 0 \\
0 & d_1 & 2d_2 & 0 & 1 & 0 \\
d_1^2 & d_1d_2 & d_2^2 & d_1 & d_2 & 1
\end{bmatrix}
\]

However, it is easy to see that \( H^T C H = C \). Therefore, the estimated ellipse will simply be a translated version of the original one. As a consequence of this, the bias and covariance of the estimated parameters \( (\alpha_1, \alpha_2, \phi_\Delta, C_1, C_2) \) are independent of \( C_1 \) and \( C_2 \) (since these quantities correspond to location parameters). Therefore, in the subsequent analysis we will assume that \( C_1 = C_2 = 0 \) without loss of generality.

A. Asymptotic Behavior of \( D^T D \)

From (16) and the assumption of having independent measurements, we have that

\[
D := \lim_{N \to \infty} \frac{1}{N} D^T D
\]

\[
\begin{bmatrix}
\bar{E} y_1^1 y_2^1 & \bar{E} y_1^1 y_2^2 & \bar{E} y_1^2 y_2^1 & \bar{E} y_1^2 y_2^2 \\
\bar{E} y_1^1 y_2^1 & \bar{E} y_1^1 y_2^2 & \bar{E} y_1^2 y_2^1 & \bar{E} y_1^2 y_2^2 \\
\bar{E} y_1^1 y_2^1 & \bar{E} y_1^1 y_2^2 & \bar{E} y_1^2 y_2^1 & \bar{E} y_1^2 y_2^2 \\
\bar{E} y_1^1 y_2^1 & \bar{E} y_1^1 y_2^2 & \bar{E} y_1^2 y_2^1 & \bar{E} y_1^2 y_2^2 \\
\bar{E} y_1^1 y_2^1 & \bar{E} y_1^1 y_2^2 & \bar{E} y_1^2 y_2^1 & \bar{E} y_1^2 y_2^2 \\
\bar{E} y_1^1 y_2^1 & \bar{E} y_1^1 y_2^2 & \bar{E} y_1^2 y_2^1 & \bar{E} y_1^2 y_2^2
\end{bmatrix}
\]
where \( \bar{E} f := \lim_{N \to \infty} N^{-1} \sum_{n=1}^{N} E \{ f(n) \} \). Here we have:

\[
\begin{align*}
\bar{E} y_1 &= \bar{E} y_2 = 0 \\
\bar{E} y_1^2 &= \frac{\alpha_2^2}{2} + \sigma_1^2 \\
\bar{E} y_2^2 &= \frac{\alpha_2^2}{2} + \sigma_2^2 \\
\bar{E} y_1 y_2 &= \frac{\alpha_1 \alpha_2 \cos(\phi \Delta)}{2} \\
\bar{E} y_1^3 &= \bar{E} y_2^3 = \bar{E} y_1^2 y_2 = \bar{E} y_1 y_2^2 = 0 \\
\bar{E} y_1^4 &= \frac{3}{8} \alpha_1^4 + 3 \sigma_1^4 + 3 \sigma_1^2 \alpha_2^2 \\
\bar{E} y_2^4 &= \frac{3}{8} \alpha_2^4 + 3 \sigma_2^4 + 3 \sigma_2^2 \alpha_1^2 \\
\bar{E} y_1^3 y_2 &= \frac{\alpha_1^3 \alpha_2}{8} \cos(\phi \Delta) \\
\bar{E} y_2^3 y_2 &= \frac{\alpha_1^3 \alpha_2}{8} \cos(\phi \Delta) \\
\bar{E} y_1^2 y_2^2 &= \frac{\alpha_1^2 \alpha_2^2}{4} \left[ 1 + \frac{1}{2} \cos(2\phi \Delta) \right] + \frac{\alpha_1^2}{2} \sigma_1^2 + \frac{\alpha_2^2}{2} \sigma_2^2 + \sigma_1^2 \sigma_2^2.
\end{align*}
\]

Notice that most of these terms depend on \( \sigma_1^2 \) or \( \sigma_2^2 \), so it can be expected that the ellipse fitting estimator will be asymptotically biased. In fact,

\[
\text{bias} \left\{ \lim_{N \to \infty} \frac{1}{N} \mathbf{D}^T \mathbf{D} \right\} =
\begin{bmatrix}
3 \sigma_1^4 + 3 \sigma_2^2 \alpha_1^2 & 0 & \beta & 0 & 0 & \sigma_1^2 \\
0 & \beta & 0 & 0 & 0 & 0 \\
\beta & 0 & 3 \sigma_2^4 + 3 \sigma_2^2 \alpha_1^2 & 0 & 0 & \sigma_2^2 \\
0 & 0 & 0 & \sigma_1^2 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_2^2 & 0 \\
\sigmoid_1 & 0 & 0 & 0 & \sigma_2^2 & 0 \\
\end{bmatrix}
\]

where

\[
\beta := \frac{\sigma_1^2}{2} \sigma_1^2 + \frac{\sigma_2^2}{2} \sigma_2^2 + \sigma_1^2 \sigma_2^2
\]

and \( \text{bias} \{ X \} := E \{ X \} - X \). The asymptotic covariance of \( \mathbf{D}^T \mathbf{D} \), which is \( \lim_{N \to \infty} N^{-1} \text{cov} (\mathbf{D}^T \mathbf{D}) \), is a 36 \times 36 matrix, composed of elements of the form

\[
\begin{align*}
\lim_{N \to \infty} N \{ E \{ \bar{E} y_i y_j \} - (E \{ y_i \}) (E \{ y_j \}) \},
\end{align*}
\]

where \( i, j, k, l \in \{ 0, 1, 2, 3, 4 \} \) and \( \bar{E} \) is the average operator, i.e.

\[
\bar{E} \{ f \} := \frac{1}{N} \sum_{n=1}^{N} f(n),
\]

and \( \text{vec}(\cdot) \) vectorizes the matrix within the parentheses [32]. Therefore, we need to compute \( E \{ y_1^4 \} \) and \( E \{ y_2^4 \} \) for \( i = 0, \ldots, 8 \). Now,

\[
\begin{align*}
E \{ y_1^4 \} &= \sum_{k=0}^{i} \left( \begin{array}{c} i \\ k \end{array} \right) \alpha_1 \cos(\omega_0 n)^{i-k} E \{ v_1^k \} \\
E \{ y_2^4 \} &= \sum_{k=0}^{i} \left( \begin{array}{c} i \\ k \end{array} \right) \alpha_2 \cos(\omega_0 n + \phi \Delta)^{i-k} E \{ v_2^k \}
\end{align*}
\]

where

\[
E \{ v_i^k(n) \} = \begin{cases}
0, & k \text{ odd } \\
1, & k \text{ even } \end{cases}
\]

\( i = 1, 2 \)

Therefore, using Weyl's equidistribution theorem [33, Chapter 3] (assuming that \( \omega_0 \) is an irrational multiple of \( 2\pi \)), (18) can be computed as

\[
\begin{align*}
\lim_{N \to \infty} N \left[ E \{ (\bar{E} y_1 y_2) (\bar{E} y_1 y_2) \} - (E \{ y_1 y_2 \}) (E \{ y_1 y_2 \}) \right] &= \frac{1}{2\pi} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left( \begin{array}{c} i+k \atop j+l \end{array} \right) (i+j) (j+l) \\
&\cdot \int_0^{\pi} \left[ \alpha_1 \cos(x) \right]^{i+k-j-l} \left[ \alpha_2 \cos(x + \phi \Delta) \right]^{j+l-q} dx \\
&\cdot E \{ v_1^i \} E \{ v_2^j \} \\
&\cdot \int_0^{\pi} \left[ \alpha_1 \cos(x) \right]^{i+k-j-l-q} \left[ \alpha_2 \cos(x + \phi \Delta) \right]^{j+l-r-s} dx \\
&\cdot E \{ v_1^i \} E \{ v_1^j \} E \{ v_2^l \} E \{ v_2^q \}
\end{align*}
\]

Therefore, the elements of \( \lim_{N \to \infty} N \text{cov}(\text{vec}(\mathbf{D}^T \mathbf{D})) \) can be computed from (20) and (19). We will not show the resulting covariance matrix, because of obvious space limitations.

**B. Asymptotic Behavior of \( \hat{\eta} \)**

Notice that \( \hat{\eta} \), as defined in (17), satisfies the Lagrangian conditions of a local minimum:

\[
(D^T D - \lambda C) \hat{\eta} = 0 \\
\hat{\eta}^T C \hat{\eta} - 1 = 0
\]

where \( \lambda \in \mathbb{R} \) is a Lagrange multiplier. We can rewrite these equations in the vectorized form

\[
(\hat{\eta}^T \otimes I) \text{vec}(D^T D) = \lambda \text{vec}(C)
\]

where \( \text{vec}(\cdot) \) denotes the Kronecker product [32]. In order to study how small perturbations of \( \text{vec}(D^T D) \) translate into small perturbations of \( \hat{\eta} \), we can consider (21) as a function \( F : \mathbb{R}^{43} \to \mathbb{R}^7 \) which maps \( \text{vec}(D^T D) \) to the null vector. In fact,

\[
F(\text{vec}(D^T D), \lambda, \hat{\eta}) = \left( \begin{array}{c} (\hat{\eta}^T \otimes I) \text{vec}(D^T D) - \lambda \text{vec}(C) \end{array} \right)
\]

This map is continuously differentiable, with derivative

\[
\frac{\partial F}{\partial [\text{vec}(D^T D) \lambda \hat{\eta}]} = \left[ \begin{array}{c} \hat{\eta}^T \otimes I & -C \hat{\eta} & D^T D - \lambda C \end{array} \right]
\]

By the implicit function theorem [34] and inversion formulae for partitioned matrices, we have the result presented in (23) (on the next page). Therefore, the asymptotic bias in \( \hat{\eta} \), for small variances \( \sigma_1^2 \) and \( \sigma_2^2 \), is given by

\[
\text{bias}(\hat{\eta}) \simeq \left\{ (D^T D - \lambda C)^{-1} \text{vec}(D^T D) (D^T D - \lambda C)^{-1} \text{vec}(C) \right\}^{-1} \eta^T C - I \cdot (D^T D - \lambda C)^{-1} (\eta^T \otimes I) \text{vec}(D^T D)
\]
\[
\frac{\partial}{\partial \text{vec}\{D^T D\}} \begin{bmatrix}
\frac{\partial \lambda}{\partial \text{vec}\{D^T D\}} \\
\frac{\partial \gamma}{\partial \text{vec}\{D^T D\}}
\end{bmatrix}
\bigg|_{\hat{\eta} = \eta}
= \begin{bmatrix}
[\eta^T C(D^T D - \lambda C)^{-1}C\eta]^{-1}\eta^T C(D^T D - \lambda C)^{-1}(\eta^T \otimes I) \\
\{(D^T D - \lambda C)^{-1}C\eta[\eta^T C(D^T D - \lambda C)^{-1}C\eta]^{-1}\eta^T C - I\}(D^T D - \lambda C)^{-1}(\eta^T \otimes I)
\end{bmatrix}
\]
\tag{23}
\]

where \(\simeq\) denotes an equality where only the dominant terms have been retained.

By the Delta method [35], the asymptotic covariance of \(\hat{\eta}\) is given by
\[
\text{cov}(\hat{\eta}) \simeq \{(D^T D - \lambda C)^{-1}C\eta[\eta^T C(D^T D - \lambda C)^{-1}C\eta]^{-1}\eta^T C - I\}. \\
(D^T D - \lambda C)^{-1}(\eta^T \otimes I)\text{cov}[\text{vec}(D^T D)](\eta \otimes I)(D^T D - \lambda C)^{-1}
\cdot \{C\eta[\eta^T C(D^T D - \lambda C)^{-1}C\eta]^{-1}\eta^T C(D^T D - \lambda C) - I\}.
\tag{24}
\]

Now, assuming that \(D := \lim_{N \to \infty} N^{-1}D^T D\) is nonsingular, we obtain (for any fixed \(\lambda\)) that
\[
\lim_{N \to \infty} N\text{cov}(\hat{\eta}) = \{(D^{-1}C\eta[\eta^T CD^{-1}C\eta]^{-1}\eta^T C - I)D^{-1}
\cdot \lim_{\varepsilon \to 0}(D + \varepsilon^{-1}C\eta\eta^T C)^{-1}\}.
\tag{25}
\]

Therefore, (24) can be written as
\[
\lim_{N \to \infty} N\text{cov}(\hat{\eta}) = \lim_{\varepsilon \to 0}(D + \varepsilon^{-1}C\eta\eta^T C)^{-1} \cdot \left\{ \lim_{N \to \infty} N^{-1}\text{cov}[\text{vec}(D^T D)] \right\} (\eta \otimes I)
\cdot \lim_{\varepsilon \to 0}(D + \varepsilon^{-1}C\eta\eta^T C)^{-1}.
\tag{26}
\]

and, as for (22),
\[
\text{bias}(\hat{\eta}) \simeq -\left\{ \lim_{N \to \infty} N^{-1}\text{vec}(D^T D) \right\} (\eta^T \otimes I)\text{bias}\left[ \lim_{N \to \infty} N^{-1}\text{vec}(D^T D) \right].
\]

C. Asymptotic Behavior of \(\hat{\vartheta}\)

Let \(\vartheta\) be given by (8). Using the Delta method, the relations between the elements in \(\eta\) and \(\vartheta\), and (25), we have that
\[
\text{cov}(\vartheta) \simeq \frac{\partial \vartheta}{\partial \eta^T} \text{cov}(\hat{\eta}) \frac{\partial \eta^T}{\partial \eta}
\]
where the gradients must be evaluated at the true value of \(\eta\), and
\[
\text{bias}(\vartheta) \simeq \frac{\partial \vartheta}{\partial \eta^T} \text{bias}(\hat{\eta}).
\]

D. High SNR Approximation

The analysis carried out in the previous subsections gives a full description of the asymptotic covariance of the ellipse fitting algorithm. However, the size of the resulting expressions is too long to provide some intuition about the qualitative behavior of the method. Therefore, in order to obtain such an intuition, we will consider a special case based on the following simplifying assumptions:

1. \(\sigma_1^2 = \sigma_2^2 = \sigma^2\).
2. \(\phi_{\Delta} = \pi/2\).
3. \(\sigma\) is small.

Under these conditions, the asymptotic covariance of the parameters \((\alpha_1, \alpha_2, \phi_{\Delta}, C_1, C_2)\) is:
\[
\text{cov}(\vartheta) \simeq \begin{bmatrix}
\frac{\sigma^2}{2N\alpha_1^2} & \frac{\sigma^2}{2N\alpha_2^2} & 0 & 0 & 0 \\
\frac{\sigma^2}{2N\alpha_1^2} & \frac{\sigma^2}{2N\alpha_2^2} & 0 & 0 & 0 \\
0 & 0 & 2SNR_{1+SNR_{2}} & 0 & 0 \\
0 & 0 & 0 & \frac{\sigma^2}{2N\alpha_1^2} & 0 \\
0 & 0 & 0 & 0 & \frac{\sigma^2}{2N\alpha_1^2}
\end{bmatrix}
\]

where the SNR was defined in (11). As it can be noted by comparing this matrix with Table I, the variances of \(\alpha_1\) and \(\alpha_2\) are within 5/4 and 3/2 times their Cramér-Rao bound, the variance of \(\phi_{\Delta}\) is twice its CRB, and the variances of \(C_1\) and \(C_2\) are within 3/2 and 2 times their CRB. This implies that the ellipse fitting algorithm is not statistically efficient.

Under the same assumptions, the bias of the parameters \((\alpha_1, \alpha_2, \phi_{\Delta}, C_1, C_2)\) is:
\[
\text{bias}(\vartheta) \simeq \begin{bmatrix}
\frac{5\alpha_2^2 - 3\alpha_1^2}{2\alpha_1\alpha_2} \sigma^2 \\
\frac{5\alpha_1^2 - 3\alpha_2^2}{2\alpha_2\alpha_1} \sigma^2 \\
0 \\
0 \\
0
\end{bmatrix}
\]
E. Results for IQ parameters

In order to determine the variance and bias of the IQ parameter estimates $G$, $Q$ and $L$ (obtained by simply replacing the parameters $(\alpha_1, \alpha_2, \phi, C_1, C_2)$ by their estimates in the definitions of such quantities), we can use the Delta method and Taylor series expansions. This gives (using the notation of Table I):

\[
\text{var}(\hat{G}) \approx \frac{2G^2}{N} \left[ \frac{1}{\text{SNR}_1} + \frac{1}{\text{SNR}_2} \right]
\]

\[
\text{var}(\hat{Q}) \approx 2 \frac{\text{SNR}_1 + \text{SNR}_2}{N \text{SNR}_1 \text{SNR}_2}
\]

\[
\text{var}(\hat{L}) \approx \frac{2L(2 + L)}{NSNR}
\]

and

\[
\text{bias}(\hat{G}) \approx \frac{4(\alpha_1^2 - \alpha_2^2)\sigma^2}{\alpha_1^4\alpha_2}
\]

\[
\text{bias}(\hat{Q}) \approx 0
\]

\[
\text{bias}(\hat{L}) \approx -\frac{L}{SNR}
\]

Note that for the statistics of $\hat{L}$ we have assumed $\alpha_1 = \alpha_2$ (but not $C_1, C_2 = 0$) in order to obtain expressions comparable to those of Table I.

The bias of $\hat{G}$ is positive if and only if $\alpha_1 > \alpha_2$, hence $\hat{G}$ is biased towards 1. This agrees with [2], where it is observed that the ellipse fitting algorithm possesses a low eccentricity bias, i.e., it tends to give ellipses which are closer to being a circle than a line.

V. SIMULATION EXAMPLE

In order to verify the theoretical results presented in the previous section, we consider a scenario presented in [20], where the performance of a nonlinear least squares method was analyzed. The setup is given by $G = 1.0$ dB, $Q = 1.0^\circ$ and $L = -40.0$ dB. The noise affecting the measurements is white and Gaussian, with $\text{SNR}_1 = \text{SNR}_2 = 74.0$ dB, which corresponds to the signal-to-noise ratio due to the use of an analog-to-digital converter of 12 bits. $N = 128$ samples are taken, and the angular frequency of the sinusoids is $\omega_0 = 0.15$ rad/s. The result of running 100000 Monte Carlo simulations is presented in Figure 1, which shows the empirical histogram of the estimates obtained using the ellipse fitting algorithm, in conjunction with their theoretical (asymptotic) distributions, and the distribution of a normal asymptotically efficient estimator (such as the one presented in [20]).

Figure 1 shows the excellent agreement between the simulation and theoretical results, even for a small sample size of 128. It is also interesting to note that the theoretical bias of the IQ parameters under this scenario is quite negligible: $|\text{bias}\{G\}| \approx -77.9$ dB and $|\text{bias}\{L\}| \approx -114.0$ dB.

VI. EXPERIMENTS

In this section the ellipse based IQ imbalance estimation scheme derived in Section III is applied on measurements obtained from five samples of the universal software radio peripheral (USRP2), all equipped with a XCVR2450 daughterboards (www.ettus.com). These daughterboards are designed around the Maxim MAX2829 BiCMOS transceiver chip (www.maxim-ic.com/datasheet/index.mvp/id/4532/t/al) which is aimed at smart-antenna/MIMO applications in the 2.4/5GHz ISM-bands using IEEE802.11b related standards. Here we investigate the scattering of the IQ parameters, $\vartheta$, using batches of only $N = 128$ samples as compared to using $N = 30000$ samples. We further investigate whether the IQ parameters, $\vartheta$, vary with the input signal frequency $\omega_0$, see (4). The variation with frequency $\omega_0$ would indicate if a frequency dependent compensation scheme is needed [37]. We further investigate the variation with respect to carrier frequency, time and across hardware units.

In a commercial implementation of the IQ parameter estimation scheme one may either perform the measurements during production tests or using on-line measurements when the node is idle. A node with two antennas could transmit the calibration signal from one of its transceivers to the other. A third possibility is to consider the IQ parameters as part of the transmission channel and estimate them in every burst [38]. Here we investigate the usefulness of the first two of these approaches by studying the variation of the IQ parameters, $\vartheta$, with base-band frequency, carrier frequency, time, and between individual XCVR2450 daughterboards. Obviously, if there are no variations in time or across units a single calibration table would suffice, and neither factory- nor online-calibration would be required. However, the results show significant variation in all dimensions except time. This indicates that a factory calibration would indeed be sufficient. However, the variation with base-band and carrier frequency shows that a table of calibration values would be needed.

A. Measurement setup and pre-processing

The receiving end of one XCVR2450 is connected to the transmitting end of another XCVR2450 by means of a cable and 25dB of extra attenuators. The signal from the transmitter XCVR2450 is generated in Matlab using a base-band sample-frequency of 25MHz. The signal is then transferred from the host PC to the USRP2 which up-samples and D/A converts it at 100MHz. The D/A converters (IQ) of the USRP2 are connected to the transmitter XCVR2450. Conversely, on the receiver side the XCVR2450 are connected to A/D converters which sample at 100MHz. The samples are then decimated to 25MHz and sent to the hosting PC by the USRP2. The USRP2 and the XCVR2450s are both sample and carrier-frequency synchronized from the same 10MHz reference. The sampling is triggered using the “time-stamp” functionality of the USRP2 UHD driver software (www.ettus.com). The driver has been slightly modified to remove any digital up- or down-conversion.

Since the transmitter as well as the receiver have DC offset and IQ imbalance, it is normally difficult to separate the effects of the transmitter and receiver. However, here we add an offset of 500kHz between the transmitter and receiver carrier (tuning) frequencies. Thus, when we generate a base-band cisol of frequency $f_{TX}$ at the transmitter, it is received at frequency...
Fig. 1: Histograms of the estimates of $G$, $Q$ and $L$, obtained using the ellipse fitting algorithm (shown as circles). The solid lines correspond to the asymptotic distribution predicted by the theory of Section IV. The dashed lines represent the asymptotic distribution of an efficient estimator. The true values are shown with a vertical dotted line.

$f_{\text{TX}} + 500$ kHz. The mirror signal of the transmitter is located at $-f_{\text{TX}} + 500$ kHz while the receiver mirror signal is located at $-f_{\text{TX}} + 500$ kHz. The DC offset of the transmitter will turn up as a cisoid of 500 kHz while the DC offset of the receiver is located at 0 Hz.

In order to test several base-band frequencies, the base-band signal is chosen as a sequence of cisoids with frequencies between $-10.5$ MHz to $9.5$ MHz with a step size of $0.5$ MHz. Each cisoid contains 30000 useful samples plus guard time for transients. When analyzing the IQ parameters of the receiver the mirror and DC component of the transmitter are first filtered. When analyzing the IQ parameters of the transmitter, the signal is first frequency translated in baseband by multiplying the received data by a cisoid with $-500$ kHz frequency, followed by filtering to remove the receiver mirror frequency and DC offset. After performing these pre-processing steps, the real and imaginary part of the estimations are used as $y_1$ and $y_2$ respectively.

**B. Small-sample performance**

In Figure 2, histograms of the $G$, $Q$ and $L$ estimates obtained from the receiver of XCVR2450 number one are shown. The results are obtained using a batch of $N = 128$ samples. Also indicated, in red, is the estimate obtained using all samples in a single batch. The results show that the small batch estimation deviates up to 0.03 dB in gain imbalance and up to 0.3 degrees in phase, at the receiver. We have found that the estimation variance decreases in inverse proportion to the number of samples, as predicted by the theoretical analysis. The number of samples needs to be selected in order to obtain the desired performance.

The phase and gain imbalance at the transmitters have been found to be consistently much smaller than at the receiver. In the following we will therefore concentrate on the receiver imbalance.

The DC offset is of less importance than the phase skew and gain imbalance because OFDM modulations generally does not use the subcarrier at the DC component.

The numbers presented here are well in-line with those stated in the data-sheet of the circuit.
C. Variation with base-band frequency $\omega_0$

Figure 3 shows the variation of the gain imbalance and phase-skew for the receiving end of all our five XCVR2450 units as a function of the base-band frequency $\omega_0$. The results show that the phase skew is clearly varying with $\omega_0$ while the gain imbalance is rather constant. On the other hand, the variation in gain imbalance is substantial among units. These results show that calibration and compensation methods need to consider variations with respect to the base-band frequency $\omega_0$ and to use individual compensations for different hardware units.

D. Variation with carrier frequency

Figure 4 shows the variation of the gain imbalance and phase-imbalance of the receiving end of all our five XCVR2450 units as a function of the carrier frequency. The base-band frequency is fixed at $-5$ MHz. The gain imbalance is varying so much that a carrier frequency dependent compensation is needed.

E. Variation with time

Figure 5 shows the variation of the gain imbalance and phase-imbalance of the receiving end of two XCVR2450 during an eighteen hour run. One measurement was done every minute. The base-band frequency is fixed at $-5$ MHz. It should be noted that there was actually a gap of 66 hours between measurements #540 and #541. The two receivers were measured almost simultaneously. At about 6 hours the measurement sequence of both receivers show a transient. Since both measurements were conducted simultaneously this transient may have a common source. The most likely cause is a rising temperature. Both receivers were located nearby a window and direct sunlight enters this window at this time. However, the magnitude of the transient is negligible (as it can be seen from the scale of the y-axis). Therefore, our measurements indicate that a calibration in production should be enough.

VII. Conclusions

In this paper we have introduced and analyzed a method for the estimation of the IQ imbalance parameters in transceivers. The method, based on an ellipse fitting technique, is simple, fast, non-iterative and relatively accurate. Our analysis has shown that the proposed method has low bias and a variance within a factor of 2 of the CRB for the $G$ and $Q$ parameters, and a factor of 4 for the $L$ parameter. These results have been verified by a numerical simulation example.

We have also applied the derived method on measurements of a contemporary BiCMOS transceiver called MAXIM 2829. From these measurements we found that the IQ imbalance varied with base-band frequency, carrier frequency, and across the five different hardware units, while the variation with time was small. In our measurements the phase skew varies up to five degrees with the base-band frequency, while the amplitude imbalance varies between 0-0.3 dB over carrier frequencies.
and across hardware units. The time variation however is only 0.004dB in amplitude and 0.06degrees in phase. This indicates that the units could either be calibrated on-line when there is no transmission (in a two antenna MIMO system one antenna could transmit a calibration signal to the other). Another alternative would be to calibrate during production in which a case a table with different carrier and base-band frequencies would be needed. However, there is no need to estimate the parameters on every burst.

REFERENCES


