Sampling, Quantization and Computational Aspects of the Quadrature Lock-In Amplifier

John Leis, Christopher Kelly
Electrical, Electronic and
Computer Engineering
University of Southern Queensland
Toowoomba, Queensland 4350
Email: {john.leis, christopher.kelly}@usq.edu.au

David Buttsworth
Mechanical and Mechatronic Engineering
University of Southern Queensland
Toowoomba, Queensland 4350
Email: david.buttsworth@usq.edu.au

Abstract—The phase-sensitive or “lock-in” amplifier is a fundamental tool in experimental physics, and is able to extract exceedingly small signals in the presence of noise. The lock-in operates on the principle of synchronous excitation of the system under test, which effectively moves the desired system response above the influence of 1/√f noise. The purpose of the paper is twofold: (i) to investigate the numerical aspects of implementation of the lock-in signal processing, particularly for computationally resource-constrained environments; and (ii) to investigate the tradeoff between A/D resolution and oversampling rates in this particular application, where a synchronous reference signal is available. We conclude that the quantization, sampling rate, and numerical precision aspects are somewhat interrelated when the best possible performance in terms of noise rejection is required.

I. INTRODUCTION

The lock-in amplifier or “lock-in” is a staple of very low-signal physical measurement processes. The fundamental approach is to make the measurand periodic in some way, thus shifting the DC signal to a known frequency and avoiding low-frequency flicker noise.

A lock-in amplifier is a phase-sensitive-detector which recovers the magnitude and phase information from a modulated signal with respect to a reference signal. The term “lock-in” amplifier is used because it locks onto that component of the output of the measurement system which is synchronous with the reference signal. Synchronous detection is obtained through the use of one or more mixers and filters. The mixer takes the incoming signal and multiplies it by the reference signal, leaving only the desired signal as a DC component together with higher-order harmonics of the original signal. The output of the mixer is filtered to remove all frequency components except the DC component containing the magnitude information. A second mixer is usually used with a quadrature (90°) reference signal, so as to obviate the need for manual phase adjustments of the reference signal. The lock-in amplifier is described in more detail in Section III.

II. PREVIOUS WORK

Recently, Aguirre et al. discussed a lock-in amplifier for portable sensing systems [1], and Son et al. proposed the use of the second harmonic rather than the more usual lowpass filtering approach [2].

Precision aspects of digital lock-in amplifiers were investigated in detail by Clarkson et al. [3], who investigated the measurement uncertainty and effects of signal to noise ratio on lock-in estimates. The work described in this paper involves a digital lock-in amplifier [4].

Andersson et al. investigated the analog-to-digital (A/D) converter resolution issue, with specific application to gas spectroscopy [5]. It was noted that gas absorption measurements are possible in the 10^{-3} – 10^{-4} concentration range, whereas modulation techniques give detection limits of 10^{-6} or better, with averaging times in the range of 1-60 seconds.

Recently, the software modelling and measurement aspects of lock-in amplifiers were investigated in detail [3], and this work provides a perspective on the signal processing aspects – in particular, the noise modelling and numerical precision requirements.

Because the signal-to-noise ratio of systems requiring the lock-in approach is quite poor, long integration times are required in order to resolve the underlying signal. This combination of large data sets and small underlying signal levels means that computational residuals have the potential to adversely affect the result. This is especially so where the computation is only able to be performed to a limited precision. Furthermore, embedded systems generally have lower precision sampling hardware, which can however be utilized at a higher sampling rate. Such oversampling approaches are quite well suited to this particular application.

We have recently discussed roundoff errors in this context, and proposed one particular type of simplification [6], as have other authors [7]. This paper extends the previous work by providing further experimental evidence and comparison to a theoretical proof. We also propose a recursive structure, and show that its performance may be substantially suboptimal unless certain precautions are taken. Finally, we also discuss the sampling aspects, and in particular the use of oversampling for the quadrature-phase lock-in amplifier.

III. SIGNAL PROCESSING FOR LOCK-IN DETECTION

As mentioned above, early approaches to lock-in amplification were purely analog ones. A simplified single-reference approach is depicted in Figure 1. For reasons of simplicity,
early implementations used a square wave reference, rather than the sinusoidal references as illustrated. This meant that the multiplication operation indicated could be replaced by a simple polarity inversion. The system under test is excited by the test signal (sine or square), and the output amplified and multiplied by a reference signal \( x_r(t) \). This reference is usually phase-shifted by an amount \( \varphi_r \), and in an experimental situation this is adjusted manually so as to maximize the output amplitude. This is obviously not desirable, and in an automatic or embedded device, this may not be possible to accomplish at all. We return to this issue shortly.

The output of the multiplier is then fed to a lowpass filter, which would ideally be a DC only filter. Thus the overall system result is that the signal of interest is translated away from DC to a higher frequency, with the modulation producing a beat effect to transform the low-level output of the system back into a DC signal. This approach avoids low-frequency noise, which is predominant in experimental situations. In particular, the so-called \( 1/f \) or flicker noise decreases in strength as a function of frequency \([8], [9]\).

To analyze the operation, we need expressions for phase-shifted sinusoidal signals multiplied by each other. Using well-known expansions for \( \cos(\alpha \pm \beta) \) and \( \sin(\alpha \pm \beta) \), we find useful relationships for our analysis in this context are

\[
\sin \alpha \sin \beta = \frac{1}{2} \left( \cos(\alpha - \beta) - \cos(\alpha + \beta) \right) \tag{1}
\]

\[
\sin \alpha \cos \beta = \frac{1}{2} \left( \sin(\alpha + \beta) + \sin(\alpha - \beta) \right) \tag{2}
\]

\[
\cos \alpha \cos \beta = \frac{1}{2} \left( \cos(\alpha - \beta) + \cos(\alpha + \beta) \right) \tag{3}
\]

Returning to the single-reference approach as depicted in Figure 1, we have

\[
x_s(t) = A_s \sin(\omega_s t + \varphi_s) \tag{4}
\]

The reference is then

\[
x_r(t) = A_r \sin(\omega_r t + \varphi_r) \tag{5}
\]

The product is

\[
x_m(t) = A_s A_r \sin(\omega_s t + \varphi_s) \sin(\omega_r t + \varphi_r) \tag{6}
\]

We now assume the system under test is linear, and thus no intermodulation products, and thus \( \omega_s = \omega_r \). Using the trigonometric expansions

\[
x_m(t) = \frac{A_s A_r}{2} \left( \cos(\varphi_s - \varphi_r) - \cos \left( 2\omega_s t + (\varphi_s + \varphi_r) \right) \right) \tag{7}
\]

Thus we have a DC component, and one at frequency \( 2\omega_s \). Assuming for the moment that we can apply a perfect lowpass filter, we have

\[
y(t) = \frac{1}{2} A_s A_r \cos(\varphi_s - \varphi_r) \tag{8}
\]

If we can adjust the phase such that \( \varphi_s = \varphi_r \), then it follows that

\[
A_s = \frac{2 x_o(t)}{A_r} \tag{9}
\]

The primary limitation of such a device is that the phase must be incrementally adjusted by some means. This may be overcome by using a phase-quadrature approach, as depicted in Figure 2. Here we have a reference signal, together with a \( 90^\circ \) phase-shifted reference. This has the advantage that we do not need to precisely (and manually) adjust the phase compensation \( \varphi_r \).

We can analyze the operation of this system as follows (we use a notation similar to that of [3], but use radian frequency for simplicity). The source signal is

\[
x_s(t) = A_s \sin(\omega_s t + \varphi_s) \tag{10}
\]

with in-phase and quadrature signals

\[
x_r(t) = A_r \sin(\omega_r t + \varphi_r) \tag{11}
\]

\[
y_r(t) = A_r \sin \left( \omega_r t + \varphi_r + \frac{\pi}{2} \right) \tag{12}
\]

\[
= A_r \cos(\omega_r t + \varphi_r) \tag{13}
\]

The product of these is

\[
x_m(t) = A_s A_r \sin(\omega_s t + \varphi_s) \sin(\omega_r t + \varphi_r) \tag{14}
\]

\[
y_m(t) = A_s A_r \sin(\omega_s t + \varphi_s) \cos(\omega_r t + \varphi_r) \tag{15}
\]

Simplifying using the trigonometric identities and again with a linear response such that \( \omega_s = \omega_r \),

\[
x_m(t) = \frac{A_s A_r}{2} \left( \cos(\varphi_s - \varphi_r) - \cos \left( 2\omega_s t + (\varphi_s + \varphi_r) \right) \right) \tag{16}
\]

\[
y_m(t) = \frac{A_s A_r}{2} \left( \sin \left( 2\omega_s t + (\varphi_s + \varphi_r) \right) + \sin (\varphi_s - \varphi_r) \right) \tag{17}
\]

After ideal lowpass filtering we have

\[
x_o(t) = \frac{1}{2} A_s A_r \cos(\varphi_s - \varphi_r) \tag{18}
\]

\[
y_o(t) = \frac{1}{2} A_s A_r \sin(\varphi_s - \varphi_r) \tag{19}
\]

Solving for the required unknowns,

\[
A_s = \frac{2}{A_r} \sqrt{x_o^2(t) + y_o^2(t)} \tag{20}
\]

\[
\varphi_s = \varphi_r + \tan^{-1} \left( \frac{y_o(t)}{x_o(t)} \right) \tag{21}
\]

Thus we have the output amplitude (which is normally what is required for this application), and the relative phase shift if desired. The amplitude is thus independent of \( \varphi_r \).

As an alternative to lowpass filtering, [2] recently suggested filtering the second harmonic \( 2\omega_s \) using a bandpass filter.
IV. PERFORMANCE LIMITATIONS

We briefly summarize some previous work on examining the performance limits of the lock-in technique. Dowell et al. examined the precision limits of waveform recovery using waveform processing techniques available at the time [10]. As noted, the goal is to maximize the signal-to-noise ratio (SNR) through appropriate signal processing strategies. It should be noted that SNR is most often stated in terms of a power ratio, but may also be stated in terms of the root-mean-square (RMS) ratio. If converted to decibels, this results in a factor of 2 difference.

Dowell et al. also noted that coherent averaging improves the SNR in proportion to the square root of the number of samples averaged. However, we point out that this improvement is in terms of the amplitude ratio rather than power. This may be derived in a manner similar to [11] as follows. For a signal with amplitude $s(t)$ and statistically independent noise samples $u(t)$, we average over $K$ samples to improve the SNR. The net signal amplitude over $K$ iterations is $A_1 + A_2 + \cdots + A_K = KA$, and the total power is proportional to the square of this. The noise sources are assumed independent, and thus the powers add independently. Thus the total noise power over $K$ observations is $U_1^2 + U_2^2 + \cdots + U_K^2 = KU^2$. The power SNR is then

$$\text{SNR}_p = \frac{P_x}{P_u}$$

$$= \frac{K^2 A^2}{KU^2}$$

$$= \frac{A^2}{U^2}$$

Thus the power SNR is improved by a factor $K$, with the amplitude SNR improved by a factor $\sqrt{K}$. Finally, Dowell et al. concludes that, considering bounds due to noise, $1/f$ noise, instrument limitations, considerable SNR improvement is possible at the expense of long averaging times.

Gillies et al. [12] subsequently established the precision limits of lock-in amplifiers based on an empirical study of real units, and in passing noted the importance of arithmetic precision. They employed averaging times longer than 100s, and concluded that the output measurement error is inversely proportional to the SNR at the input. Subsequently, [13] employed the experimental results of [12] and showed that such results are predictable on a theoretical basis. The RMS noise is proportional to the square-root of the bandwidth. Defining the effective or equivalent-noise bandwidth

$$\Delta f = \frac{1}{2\pi} \int_0^\infty \left| \frac{G(\omega)}{G_{\text{max}}(\omega)} \right|^2 d\omega$$

a first-order Butterworth-characteristic filter defined by

$$|G(\omega)| = \frac{1}{\sqrt{1 + (\omega / \omega_c)^2}}$$
has an ENBW of $\Delta f = \frac{\pi}{2f_c}$. Thus, the problem may be framed from the point of view of bandwidth reduction. However, the limiting case is the fact that a lowpass filter is employed, thus providing an upper bound on the improvement possible. The long time constant employed (125s) gave an effective bandwidth (assuming a 1st-order filter) of 0.002 Hz. For 5 MHz measurement bandwidth, the RMS of the input noise is reduced by a factor of $\sqrt{\frac{BW_{in}}{BW_{out}}}$. Finally, since the bandwidth $BW_{out}$ is proportional to the time constant of the input integration, the linear dependency noted experimentally reduces to this ratio.

V. SNR Performance Limits

Figure 4 shows some experimentally derived performance limits, in terms of minimum detectable SNR and the corresponding block size for integration. We have arbitrarily used an error of 5% of the true value as being acceptable – this value may need to be adjusted down in some applications.

Figure 4 shows a 10 dB increase in the sensitivity of the lock-in with a tenfold increase of the block size. This relationship is directly related to the quality of the low pass filter used in the lock-in amplifier. As the number of samples increases the bandwidth of the low pass filter decreases, thus the lock-in rejects more noise.

VI. Accuracy of Recursive Computations

The FIR filter equation is

$$y(n) = \sum_{k=0}^{N-1} h_k x(n-k)$$  \hspace{1cm} (27)$$

$$= h_0 x(n) + h_1 x(n-1) + h_2 x(n-2) + \cdots$$  \hspace{1cm} (28)$$

Bomar [14] derived the error variance for the standard FIR filter as

$$\sigma_e^2 = \left[ Nh_0^2 + \sum_{k=1}^{N-1} (N+1-k)h_k^2 \right] \sigma_x^2 \sigma_e^2$$  \hspace{1cm} (29)$$

where $\sigma_x^2$ is the variance of the input random variable, and $B$ is the number of significand bits in the fractional representation ($B = 24$ for IEEE754 standard single-precision). The growth of the summation error is proportional to $N$, which clearly may be reduced in magnitude if the order is such that larger $(N+1-k)$ is associated with smaller $k$. That is, if smaller $h_k$'s are associated with smaller $k$. This equation demonstrates the basis of the often-held assertion that it is better to add lists of numbers in ascending order. If $h_k = 1$, the summation may be reduced to $\frac{1}{2}(N-1)(N+2)$

For $B$ fraction bits then the step size is

$$\Delta = 2^{-B}$$  \hspace{1cm} (30)$$

For floating-point arithmetic adhering to IEEE754 standard, $B = 24$ not 23 because of the 1 to the left of the binary point. The floating-point quantization error is $-\frac{\Delta}{2} < \varepsilon < \frac{\Delta}{2}$ with mean $\frac{\Delta}{2}$ with mean $\frac{\Delta}{2}$, so the ratio of the relative error to the mean is $\frac{\Delta}{m}$, whose variance is

$$\sigma_e^2 = \frac{1}{1/2} \int_{1/2}^{1} \left[ \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \left( \frac{\varepsilon}{m} \right)^2 \varepsilon \right] \varepsilon dm$$  \hspace{1cm} (31)$$

$$= \frac{\Delta^2}{6}$$  \hspace{1cm} (32)$$

Figure 5 shows a comparison of experimental results with theoretical roundoff error as predicted by (32). For comparison, the summation of samples taken in a pairwise fashion, and samples summed in sorted order, are also shown. Taking values in order of magnitude is able to reduce the error due to rounding, but the improvement is not especially significant. Additionally, this method would require a numerical sort stage, which is somewhat time-consuming. However, taking values in pairs and producing intermediate sums clearly reduces the error for all block sizes. Summation methods for long vectors of floating-point numbers have been discussed extensively in the mathematical literature on numerical analysis [15], [16].

VII. Comparison of Arithmetic Limitations

We now provide some results in the context of the lock-in amplifier. We define the total integration block length as $N$, and the number of samples in each sine wave cycle as $M$. Thus the phase angle per sample is $2\pi/M$. In the following results, we assume that the true and correct result is obtained by iteratively computing the multiply-integrate steps using double precision arithmetic, and that all results are likewise stored in double precision format.

The averaging stage expressed as an FIR filter is

$$y(n) = \frac{1}{N} \left[ x(n) + x(n-1) + x(n-2) + \cdots \right.$$

$$\cdots + x(n-N+1) \right]$$  \hspace{1cm} (33)$$

$$= \frac{1}{N} \sum_{m=0}^{N-1} x(n-m)$$  \hspace{1cm} (34)$$

At each time step $n$, the majority (actually, $N-1$) of the terms have already been summed. All that remains is to discard
Fig. 5. Floating-point summation errors, comparing the theoretical value with experimentally-generated results. The pairwise summation approach, discussed previously, is also shown.

(subtract out) the oldest term \(x(n-N)\), and incrementally add the newest sample \(x(n)\). Mathematically, we can derive this by forming the sums for \(y(n)\) and \(y(n-1)\), and subtracting term-wise, to give

\[
y(n) = y(n-1) + \frac{1}{N} \left[ x(n) - x(n-N) \right] \tag{35}
\]

with \(z\) transform

\[
\frac{Y(z)}{X(z)} = \frac{1 - z^{-N}}{1 - z^{-1}} \tag{36}
\]

We may combine the sine/cosine multiply stage with the averaging, using

\[
x(n) = x_r(n) x_m(n) \tag{37}
\]

It is possible to cast this in a similar recursive fashion, since the impulse response is symmetric without decay.

A recursive multiply-integrate step may be performed using double precision, with results saved as single precision. This saves memory, but requires a double-precision accumulator. Also, only one cycle of sine is required, and thus we may compute the reference in a modulo-M fashion

\[
x(n) = x_r([n \mod M]) x_m(n) \tag{38}
\]

For comparison, we show in the following the single-precision iterative approach. This is identical to the double-precision iterative approach, but performs all calculations to single-precision only. Finally, we show the error for the single-precision recursive calculation.

Figure 6 shows the performance for three approaches using a small block size \((N = 1000)\). The double-precision recursive approach presents an error distribution from the ideal, although the error is quite small (of the order of \(10^{-10}\)). The recursive single-precision approach has errors of the order of \(10^{-3}\), which may be intrusive in many applications. Using the same precision but an iterative (rather than recursive) computation reduces the average error somewhat.

Next, Figure 7 extends this to a larger block size, keeping the same relative number of samples per sinusoidal cycle. The single-precision approach when iterated produces errors of the order of 0.5\%, whereas the recursive approach has a substantially wider spread, to 10\% and beyond.

Finally, Figure 8 employs the same block size, but a much larger number of samples per cycle (400). In this case, the recursive algorithm operating at a single-precision produces results which would be deemed unacceptable.
VIII. OVERSAMPLING AND QUANTIZATION

We now investigate the effects of quantizer resolution and sampling rate. A well-known result is that each bit of additional precision increases the SNR by 6 dB, or

$$\text{SNR (dB)} \propto 6N$$  \hspace{1cm} (39)

With an oversampling ratio of $R$, defined as

$$R = \frac{f_{so}}{f_s}$$  \hspace{1cm} (40)

the noise is spread across the entire band [17, Appendix A]. Using a quantization step size $\Delta$

$$\Delta = \frac{V_{\text{max}}}{2^{N_{eq}}}$$  \hspace{1cm} (41)

the quantizing noise is assumed to be uniformly distributed between $\pm \Delta/2$, then

$$\sigma_e^2 = \frac{\Delta^2}{12}$$  \hspace{1cm} (42)

So for oversampling by a factor of $R$,

$$\sigma_e^2 = \frac{\Delta^2}{12R}$$  \hspace{1cm} (43)

Let the maximum range be

$$V_{\text{max}} = 4\sigma_x$$  \hspace{1cm} (44)

Defining $N_{eq}$ to be the equivalent number of bits (ENOB), the SNR is

$$\text{SNR} = \frac{\sigma_x^2}{\sigma_e^2}$$  \hspace{1cm} (45)

$$= \frac{\sigma_x^2}{\frac{1}{12R} \left( \frac{V_{\text{max}}}{2^{N_{eq}}} \right)}$$  \hspace{1cm} (46)

$$= \frac{3}{4}R2^{2N_{eq}}$$  \hspace{1cm} (47)

$$= \frac{3}{4}R4^{N_{eq}}$$  \hspace{1cm} (48)

The result is that each doubling of the sampling frequency (doubling $R$) lowers the noise by 3 dB [17, Appendix A]. Also, each doubling of the sampling frequency reduces the number of bits $N_{eq}$ by $\frac{1}{2}$. Equivalently, doubling the sampling frequency increases the ENOB by 0.5. Thus,

$$f_{so} = 2^{2N_{eq}}f_s$$  \hspace{1cm} (49)

$$= 4^{N_{eq}}f_s$$  \hspace{1cm} (50)

where $f_s$ is the original sampling frequency, $f_{so}$ is the oversampling frequency, and $N_{eq}$ is the ENOB.

Figure 9 shows that for SNR levels of $-40$ dB and below, there is no appreciable difference between using an 8 bit and a 14 bit A/D converter, as below 40 dB SNR the mean-squared error is directly proportional to the power of the noise. This means that a lower-complexity A/D converter may be employed, provided oversampling is also used. The assumed theoretical noise floor is that of the quantization noise, since from Figure 9 the mean-square error is around $10^{-5}$ for 8 bit resolution. However, this is not the case for the detectable SNR, as quantization noise exceeds the input noise just after 0 dB SNR.

We point out the connection with the long-established process of dithering before quantization. Dithering is a process used in sampling where another signal is added with the input signal before the A/D converter to boost the input signal so that it exceeds the quantizer step size, thus making it visible to the A/D converter. In this way, dithering can effectively increase the resolution of the A/D converter. Because the dithering signal is known, it can be removed in software leaving behind the input signal. This does not apply, however, in the case of low-level signal measurements.
IX. CONCLUSIONS

We have shown that it is possible to exploit the additional noise in lock-in applications to our advantage, by using a lower resolution A/D converter in conjunction with oversampling. This is because the additional noise effectively acts as a dither signal for larger step sizes. However, using a large oversampling ratio $R$ results in the requirement for a large integration block $N$. This means that greater care must be taken in calculating the sine-cosine products and integrating them, because the rounding effects are cumulative. More specifically, single-precision calculations can cause large roundoff errors to accumulate. Further, the computationally-efficient recursive algorithm causes roundoff errors to accumulate at a larger rate. Thus the fundamental tradeoff involves sampling accuracy, sampling rate, memory buffer requirements, and computational accuracy.

REFERENCES