A DOUBLE DEMODULATION ALGORITHM FOR NARROW SPECTRAL FEATURES

Enrico Ragaini
Dipartimento di Elettrotecnica, Politecnico di Milano

Ronald F. Woodman
Radio Observatorio de Jicamarca

Abstract

This paper proposes a technique for processing signals in which the relevant feature consists of the shape and position of narrow Doppler spectral lines. The proposed technique is the digital version of a quadrature demodulator, and its purpose is to demodulate a selected spectral line allowing to study its envelope as a slowly varying process.
First, we review the usual quadrature demodulation and coherent integration techniques, which are of common use in Doppler radar.
The principle of double demodulation is to apply a frequency shift to signals to bring the selected spectral line near zero frequency, and then to use a lowpass filter to discard the remaining spectral components. This principles are exposed in a mathematical way, using complex signals and exponentials.
We then show how this idea is translated into an efficient algorithm, employing real multiplications and accumulators, easy to implement on a radar signal processing system.
After this discussion, a possible simplification of the algorithm is discussed, in which sinusoids used for demodulation are replaced by square waves, so that multiplication reduces to a synchronous sign reversal. Possible distorsions due to this simplified algorithm are mentioned.
In the final part of the paper, some possible implementations of double demodulation are proposed, and the advantages of each are discussed.

1. Introduction.

Doppler radar experiments are based on spectral analysis. In fact, information is obtained from data by evaluation of their frequency content. For example, in a wind profiling experiment, physical quantities are obtained by evaluation of the Doppler shift and width of the frequency spectrum of clear air echoes.
In a generic Doppler radar experiment, the measurement system produces a set of frequency spectra of the received signal. A narrow spectral band centered on the frequency of the transmitted signal is observed, so that frequency offsets (Doppler shifts) can be appreciated. Typical observed bandwidths range from some hertz to some hundreds of hertz.
In some experiments, the relevant part of the measured spectrum consists of one or more narrow spectral lines. Linewidths of fractions of a hertz are common: typical examples are sea echoes in HF or VHF measurements, and acoustic lines in radio-acoustic sounding systems. In these cases, the physical measurement is based on the evaluation of the shape and position of such lines. Thus, a high frequency resolution is required in spectral evaluation in order to obtain significant results.
Signal processing systems which are commonly used in MST radar systems usually employ FFT-based algorithms to evaluate frequency spectra. With these algorithms, analyzing a narrow spectral line in detail is often impossible, because the same frequency resolution has to be applied to the whole observed band, so that a large number of spectral points have to be used. This can easily become an overwhelming requirement for the signal processing system,

especially when many range gates have to be processed in parallel (i.e., many spectra have to be calculated at once).
The following set of values can give an idea of the problem: a RASS system may observe a Doppler spectrum from -300 Hz to +300 Hz. Suppose a spectral line 0.1 Hz wide lies between +100 Hz and +101 Hz. To observe the spectrum with a 0.01 Hz resolution, an FFT algorithm would produce a 60000-points spectrum. Note that only 100 of the 60000 values are significant.
In the following, we propose an easily implementable processing technique which would allow observation of narrow spectral bands without requiring extensive calculations.

2. Quadrature demodulation and coherent integration

2.1. Quadrature demodulation

Let us briefly review the usual processing employed for Doppler radar signals. To simplify the discussion, we consider observation of only one height (a single range gate), supposing that the same processing algorithm is used for all observed heights.
The signal received by the antenna is a real, narrowband (or quasi-sinusoidal) process. Its frequency spectrum is confined to a narrow band centered on a carrier frequency, namely the frequency of the transmitted signal. This is a consequence of the physical processes involved, because the received signal can show a range of Doppler shifts, positive or negative, with respect to the transmitted frequency.
A narrowband signal can be expressed as the product of a slowly varying signal and a fixed frequency carrier. The received signal *r(t)* can therefore be written as:

$$r(t) = s(t)e^{j\Omega t} + s^*(t)e^{-j\Omega t}$$

where Ω is the frequency of the transmitted signal. The second term exists because *r(t)* is real. We can say that *r(t)* is an amplitude-modulated signal and Ω is the carrier frequency. *s(t)* is the baseband signal. The demodulation process consists of obtaining *s(t)* from *r(t)*.
The following facts are worth remarking:

- the Fourier transform of *r(t)*, for positive frequencies, is obtained from that of *s(t)* by a translation Ω on the frequency axis. A symmetric replica of the spectrum is present at frequency $-\Omega$. If the transform of *s(t)* is $S(\omega)$, *r(t)* transforms to $S(\omega-\Omega)+S^*(\Omega-\omega)$
- all experimental information contained in *r(t)* can be recovered from *s(t)*, because the carrier is simply a sinusoidal signal of known frequency; yet, because *s(t)* is slowly varying (i.e., its spectrum is limited to a narrow band around zero frequency) it is much less demanding for a signal processing system to analyze *s(t)* than *r(t)*;
- the Doppler spectrum of the received signal is simply the spectrum of *s(t)*. Positive and negative Doppler shifts are reproduced as positive and negative frequencies in *s(t)*.
- *s(t)* is a complex signal, because the Doppler spectrum is not necesarily symmetric.

To recover *s(t)* from *r(t)*, a quadrature demodulator is used: *r(t)* is multiplied by $e^{-j\Omega t}$ and then lowpass-filtered. Multiplication yields:

$$r(t)e^{-j\Omega t} = s(t)e^{j\Omega t}e^{-j\Omega t} + s^*(t)e^{-j\Omega t}e^{-j\Omega t} = s(t) + s^*(t)e^{-j2\Omega t}$$

The second term, a narrowband signal at frequencies around 2Ω, is filtered out and *s(t)* results.

As *s(t)* is a complex signal, it can be expressed as:

$$s(t) = a(t) + jb(t)$$

$s(t)$ actually consists of the two real signals $a(t)$ and $b(t)$. To show how a quadrature demodulator retrieves them from $r(t)$, let us write the above relations in terms of their real components:

$$r(t) = Re[s(t)e^{j\Omega t}] = Re[(a(t)+jb(t))(cos\Omega t + jsin\Omega t)] = a(t)cos\Omega t - b(t)sin\Omega t$$

$$r(t)e^{-j\Omega t} = (a(t)cos\Omega t - b(t)sin\Omega t)(cos\Omega t - jsin\Omega t) =$$
$$= (a(t)cos\Omega t - b(t)sin\Omega t)cos\Omega t - j(a(t)cos\Omega t - b(t)sin\Omega t)sin\Omega t$$

The two real signals corresponding to the real and imaginary parts of the product are thus obtained by multiplying $r(t)$ by $cos\Omega t$ and $-sin\Omega t$, respectively. The results are:

$$(a(t)cos\Omega t - b(t)sin\Omega t)cos\Omega t = a(t)cos^2\Omega t - b(t)sin\Omega t\, cos\Omega t =$$
$$= \tfrac{1}{2}a(t) + \tfrac{1}{2}b(t)cos2\Omega t - \tfrac{1}{2}b(t)sin2\Omega t$$

for the real component, and

$$-(a(t)cos\Omega t - b(t)sin\Omega t)sin\Omega t = b(t)sin^2\Omega t - a(t)cos\Omega t\, sin\Omega t =$$
$$= \tfrac{1}{2}b(t) - \tfrac{1}{2}b(t)cos2\Omega t - \tfrac{1}{2}a(t)sin2\Omega t$$

for the imaginary.

Lowpass filtering cancels all the sinusoidal components, and $a(t)$ and $b(t)$ are obtained.

In fact, a quadrature demodulator works by mixing (multiplying) $r(t)$ by two sinusoids of the frequency of the local oscillator and 90° out of phase, and lowpass-filtering the resulting signals, which are then called the "in phase" and "in quadrature" components of the signal. Multiplications and filtering are implemented by analog circuits. The two output signals are then sampled and used as a complex time series for spectral analysis.

Quadrature demodulation is currently used in all Doppler processing systems.

2.2. Coherent integration as a filtering process

The sampling rate used for $a(t)$ and $b(t)$ is often much faster than the bandwidth of interest, because one sample is taken for each transmitted RF pulse. In VHF wind profiling, for example, sampling rate can be 1 kHz for a 10 Hz signal bandwidth.

In these cases, the first processing step is a digital lowpass filtering. This is usually implemented by means of time-domain averaging: blocks of successive samples are summed together, and the results are used as a new time series which is then frequency-analyzed. As the real and imaginary parts of the samples are accumulated separately, phase coherence is maintained. Time averaging is also known as coherent integration.

Time averaging can be described as the following input-output transformation: the input process is sampled at a rate f_s (sampling interval is $t_s = \frac{1}{f_s}$); one output sample is generated every N input samples, equal to the sum of the last N input samples. The output sample rate is $\frac{f_s}{N}$, so that frequencies between $\frac{-f_s}{2N}$ and $\frac{f_s}{2N}$ are observed.

An equivalent way to represent coherent integration is by a boxcar filtering (using a digital filter with rectangular impulse response N samples long) followed by a decimation which takes one sample every N. Consequences are:

- the rectangular filter is a well known lowpass device, with a transfer function $\frac{\sin \pi f N t_s}{\sin \pi f t_s}$: it is interesting to note that the first transmission zero is located at $\frac{f_s}{N}$, and the following ones at multiples of this value.
- decimation means resampling the signal at a rate N times lower, hence the observed frequency band is reduced N times, namely to the band between $\pm\frac{1}{2Nf_s}$. Frequencies outside this band are aliased.

The effect of aliasing is reduced at frequencies near zero because all frequencies that would be aliased into zero frequency are transmission zeroes of the filter.

The aliasing effect we just mentioned is not a problem when the following assumptions hold, which is the case of clear air radar experiments:

- other spectral features, if present, are far removed from the observed band, so that they are attenuated by the lowpass filter enough to be neglected;
- additive noise has a white spectrum, i.e., noise is completely uncorrelated from sample to sample. In this case, noise power in the output is multiplied by N, while signal power is proportional to N^2.

Coherent integration is usually implemented by a complex accumulator/adder. The accumulator initially contains zero. As input signal is sampled, each sample value is added to the accumulator, which contains the running sum. After N add-and-store operations, the accumulator contains the sum of the last N samples. This value is output, the accumulator is reset to zero and the process starts over again. The adder and accumulator separately process real and imaginary components.

We can consider demodulation and coherent integration together as a filtering process: the signal is first shifted down in frequency, so that the relevant frequency band lies around zero frequency. This spectral band is then selected by means of a lowpass filter. The frequency shift is equal to the local oscillator frequency, while the lowpass filter bandwidth is determined by the number of coherent integrations.

2.3. Total Observation Time

When an FFT algorithm with M points is used for spectral analysis, frequency resolution is:

$\delta f = \frac{f_e}{M}$, where f_e is the effective sampling rate. This is because M spectral points (bins) are computed, spanning a frequency band from $-\frac{f_e}{2}$ to $\frac{f_e}{2}$. A well known relationship exists between δf and the total observation time T:

$$T = \frac{M}{f_e} = \frac{1}{\delta f}$$

The result is that the minimum observation time required in order to measure a spectrum with frequency resolution δf is equal to the inverse of δf.

It is important to remark that this relationship is not affected by coherent integration. In fact, when an integration factor of N is used, the effective sampling rate is $f_e = \frac{fs}{N}$, the observed bandwidth is $\frac{fs}{N}$, and with M points a frequency resolution $\delta f = \frac{fs}{MN}$ is obtained. The total time required to obtain M data is again $T = \frac{M}{f_e} = \frac{MN}{fs} = \frac{1}{\delta f}$.

Thus, coherent integration does not reduce observation time. On the contrary, it reduces the required number of spectral points (from *MN* to *M*) because only the relevant part of the spectrum, namely the lowest frequencies, is observed. As the sampling rate is reduced by a factor of N, input rate to the spectrum analyzer is correspondingly reduced.

3. Double demodulation concept.

In the case of a narrow spectral line, the output of the quadrature demodulator is in turn a narrowband signal. This time, the term narrowband is referred to the frequency band under observation, so we mean that the signal *s(t)* is confined to a frequency band much smaller than the sampling rate, centered on some frequency ω. Also we expect that the signal bandwidth be much smaller than ω itself.

The received signal *r(t)* thus covers a narrow frequency band around a frequency Ω+ω. So, we can write *r(t)* and the demodulated signal *s(t)* respectively as:

$$r(t) = x(t)e^{j(\Omega+\omega)t} + x^*(t)e^{-j(\Omega+\omega)t}$$

$$s(t) = r(t)e^{-j\Omega t} = x(t)e^{j\omega t} + x^*(t)e^{-j(2\Omega+\omega)t}$$

And after lowpass filtering only the first term of *s(t)* remains:

$$s(t) = x(t)e^{j\omega t}$$

As the bandwidth of *x(t)* is much narrower than ω, *x(t)* is a much "slower" process than *s(t)*. In fact, the maximum frequencies that appear in *s(t)* are of the order of ω,.while the spectrum of *x(t)* is only as wide as the spectral line itself.

If ω is known, we can apply a second demodulation process to *s(t)* and recover *x(t)*. This can be done by a quadrature demodulation (i.e., multiplication by a complex exponential) followed by a coherent integration, which works as a lowpass filter. Both steps would be performed by digital circuits, as *s(t)* is available to the signal processing system in the form of a complex time series.

The first step consists in multiplication by a demodulating factor, namely a complex exponential at frequency -ω. The spectral line at ω is translated at zero frequency:

$$s(t)e^{-j\omega t} = x(t)e^{j\omega t}e^{-j\omega t} = x(t)$$

In the frequency domain, multiplication has the effect of "demodulating" *s(t)*, i.e., to shift all its spectrum down in frequency by an amount ω. As the only relevant part of the spectrum is now around zero frequency, a lowpass filter can be used.

In particular, time averaging (coherent integration) can be used. As discussed above, its effect is lowpass filtering followed by spectral folding. Of course, any other lowpass filter could be used as well.

In summary, double demodulation consists of a multiplication followed by coherent integration. A slowly varying complex signal is obtained, the spectrum of which reproduces the shape of the line under observation.

Practical realization of this algorithm starts with the generation of a series of samples of $e^{-j\omega t}$ corresponding to the sampling instants of *s(t)*. A new time series *x(t)* is then obtained as the product:

$x(nts) = s(nts)e^{-j\omega nts}$.

The resulting samples would then be coherently integrated, e.g. by an adder/accumulator as explained above.

The advantage of such an algorithm is that spectral evaluation is performed only on a restricted frequency band, so that a much smaller number of points can be used.

As a rather extreme example, suppose we want to observe, with 0.01 Hz resolution, a spectral band 1 Hz wide centered at 100 Hz. Let the sampling frequency be 1 kHz. With a simple FFT on samples, the spectrum from - 500 Hz to 500 Hz should be calculated with the required resolution, so that more than 100000 points should be used for the FFT. With simple coherent integration (N = 4) we can reduce the effective sampling rate to 250 Hz, and observe a band from - 125 Hz to 125 Hz. In order to have 0.01 Hz resolution, 25000 points would be used. With the proposed algorithm, we would multiply by a complex exponential of frequency -100 Hz and then integrate coherently 1000 times and calculate the spectrum between - 0.5 Hz and 0.5 Hz, which only requires 100 points at the required spectral resolution.

A further advantage of double demodulation is that the signal processor which calculates the FFTs receives data at a much slower rate due to coherent integration, hence more time is available for computation.

The following points should be remarked:

- the total observation time required to achieve a given spectral resolution is not affected by the algorithm; in fact, the final sampling rate, $\frac{1}{Nts}$, depends on the number of coherent integrations, but not on ω. If M samples are taken, frequency resolution is $\delta f = \frac{1}{MNts}$, and total observation time is again $T = MNfs = \frac{1}{\delta f}$
- the same algorithm can of course be applied for demodulating a spectral line at -ω (i.e., a negative Doppler shift), by using $e^{j\omega t}$ as the demodulating factor;
- no distortion is introduced by the multiplication, because a pure complex exponential is used.

If two Doppler spectral lines are present at frequencies ω and -ω, the algorithm can be used to select and examine each of the two. Yet, care must be taken to avoid aliasing. If we want to observe the line at +ω, the demodulation factor is $e^{-j\omega t}$. The band around -ω is shifted to -2ω: the lowpass filter should efficiently attenuate such a frequency to avoid significant aliasing.

Let us look in more detail at the complex multiplication involved in demodulation. As opposite to the quadrature demodulators used in receivers, the input signal is already complex, so the real and imaginary parts of *x(t)* have a somewhat more complicated expression.

We can express *x(t)* as

$$x(t) = c(t) + j\,d(t)$$

and writing the demodulation process in terms of real signals we obtain:

$$x(t) = s(t)e^{-j\omega t}$$

$$c(t) + jd(t) = (a(t) + jb(t))(\cos\omega t - j\sin\omega t) =$$
$$= a(t)\cos\omega t + b(t)\sin\omega t - ja(t)\sin\omega t + jb(t)\cos\omega t$$

so that the two components of *x(t)* are calculated as:

$$c(t) = a(t)\cos\omega t + b(t)\sin\omega t$$
$$d(t) = b(t)\cos\omega t - a(t)\sin\omega t$$

We observe that 4 real multiplications and 2 real sums are required for each sample if the above relations are used for implementation. A more efficient implementation can be suggested by the observation that samples of *c(t)* and *d(t)* are going to be accumulated in the coherent integration stage. Hence, the same result can be obtained by separately accumulating the four terms and summing them together afterwards. In formulae:

$$\sum[a(t)\cos\omega t + b(t)\sin\omega t] = \sum a(t)\cos\omega t + \sum b(t)\sin\omega t$$
$$\sum[b(t)\cos\omega t - a(t)\sin\omega t] = \sum b(t)\cos\omega t - \sum a(t)\sin\omega t$$

where we indicated the integration operator simply by Σ.

Implementation would then consist of calculating 4 time series by real multiplication:

$$A(t) = a(t)\cos\omega t$$
$$B(t) = a(t)\sin\omega t$$
$$C(t) = b(t)\cos\omega t$$
$$D(t) = b(t)\sin\omega t$$

and integrating (accumulating) them separately the required number of times. The components of the complex signal to process are then:

$$c(t) = \Sigma A(t) + \Sigma D(t)$$
$$d(t) = \Sigma C(t) - \Sigma B(t)$$

A further advantage of the above approach is that baseband signals for spectral lines at $+\omega$ and $-\omega$ can both be calculated from *A(t), B(t), C(t) and D(t)*. In fact, suppose two spectral lines are present in the signal:

$$s(t) = x(t)e^{j\omega t} + y(t)e^{-j\omega t}$$

in which

$$s(t) = a(t) + j\,b(t)$$
$$x(t) = c(t) + j\,d(t)$$
$$y(t) = c'(t) + j\,d'(t)$$

Substituting and separating real and imaginary parts yields:

$$a(t) + jb(t) = (c(t) + jd(t))(\cos\omega t + j\sin\omega t) + (c'(t) + jd'(t))(\cos\omega t - j\sin\omega t)$$
$$a(t) = c(t)\cos\omega t - d(t)\sin\omega t + c'(t)\cos\omega t + d'(t)\sin\omega t$$
$$b(t) = d(t)\cos\omega t + c(t)\sin\omega t + d'(t)\cos\omega t - c'(t)\sin\omega t$$

Multiplication of *a(t)* and *b(t)* by *cos ωt* and *sin ωt*, after some algebra, yields:

$$A(t) = \tfrac{1}{2}[c(t) + c'(t)] + \tfrac{1}{2}[c(t) + c'(t)]\cos 2\omega t + \tfrac{1}{2}[d'(t) - d(t)]\sin 2\omega t$$
$$B(t) = \tfrac{1}{2}[d'(t) - d(t)] + \tfrac{1}{2}[d(t) - d'(t)]\cos 2\omega t + \tfrac{1}{2}[c(t) + c'(t)]\sin 2\omega t$$
$$C(t) = \tfrac{1}{2}[d(t) + d'(t)] + \tfrac{1}{2}[d(t) + d'(t)]\cos 2\omega t + \tfrac{1}{2}[c(t) - c'(t)]\sin 2\omega t$$
$$D(t) = \tfrac{1}{2}[c(t) - c'(t)] + \tfrac{1}{2}[c'(t) - c(t)]\cos 2\omega t + \tfrac{1}{2}[d(t) + d'(t)]\sin 2\omega t$$

The above signals contain:

- slowly-varying signals, namely $c(t)$, $d(t)$, $c'(t)$, $d'(t)$, with constant coefficients;
- signals in the frequency band around 2ω, which appear in the form of products of the above by $\cos 2\omega t$ or $\sin 2\omega t$.

We assume that coherent integration filters out the second class of signals, so that only low frequencies remain. Expressions simplify to:

$$\Sigma A(t)=\frac{1}{2}[c(t)+c'(t)]$$

$$\Sigma B(t)=\frac{1}{2}[d'(t)-d(t)]$$

$$\Sigma C(t)=\frac{1}{2}[d(t)+d'(t)]$$

$$\Sigma D(t)=\frac{1}{2}[c(t)-c'(t)]$$

And the components of each of the two narrowband signals can be retrieved:

$$\Sigma A(t)+\Sigma D(t)=c(t)$$
$$\Sigma C(t)-\Sigma B(t)=d(t)$$
$$\Sigma A(t)-\Sigma D(t)=c'(t)$$
$$\Sigma C(t)+\Sigma B(t)=d'(t)$$

The complex time series $x(t)$ and $y(t)$ can then be separately processed to evaluate the spectra of $+\omega$ and $-\omega$ bands.

It is interesting to note that the processing system can be designed to store also the initial $s(t)$ values: if enough processing power is available, a third narrow band, around zero frequency, can also be observed in parallel to the other two.

4. Double demodulation by sign reversal.

The main difficulty in implementing the above technique in a real-time processing system is the presence of real multiplications. Values for the sinusoids would have to be calculated or extracted from a table, and data would have to be multiplied and accumulated in real-time. The processing system may be too slow to perform these operations in real time, especially when many heights have to be processed in parallel.

If available processing power doesn't allow an implementation of the complete algorithm with sine wave multiplication, a simplified version of it can be used, in which sine waves are replaced by square waves. Such simplification is obtained at the cost of some distortion of the signal, as we are now going to discuss.

The key point is the following: multiplication by a sine wave is approximated by an alternated sign reversal operation. The signal to be demodulated is multiplied by +1 or -1 according to the sign of the demodulating sine wave. The process is performed in parallel for the $\cos \omega t$ and $\sin \omega t$ components.

In other words, the $\sin \omega t$ and $\cos \omega t$ functions are replaced by a couple of square waves 90° out of phase. The demodulation and integration process is reduced to adding data samples into separate accumulators, reversing their sign when the corresponding square wave is -1 and

leaving them unaffected when it is +1. These operations are clearly of faster execution than floating point multiplications.
After such sign reversal, coherent integration, same as explained above, would be used. We are going to consider separately the effects of such operations on spectral lines, if present, and on white noise.

Consider the complex signal defined as follows:

$$h(t) = Sq(\cos \omega t) + j\, Sq(\sin \omega t)$$

where $Sq(\,)$ is a modified signum function, which only takes values +1 and -1:

$$Sq(x) = 1 \quad \text{for } x \geq 0$$
$$Sq(x) = -1 \quad \text{for } x < 0$$

$h(t)$ is the "square wave" version of a complex exponential, in which the real and imaginary part are square waves 90° out of phase. We call $h(t)$ a "complex square wave" of frequency ω, and we are going to use it as a demodulating signal.
Some calculations show that h(t) can be expressed as a Fourier series, in which many of the harmonics are missing. We can write the nonzero terms as the series:

$$h(t) = \sum_{n=-\infty}^{+\infty} \frac{8}{4n+1} e^{j(4n+1)\omega t}$$

The spectrum is thus composed by a fundamental frequency ω and by harmonics at frequencies $\omega \pm 4n\omega$. Therefore, harmonics are present at 5, 9, 13 ... and at -3, -7, -11 ... times the fundamental frequency. In particular, no harmonic is present at $-\omega$, as we expect because $h(t)$ is the result of a positive fundamental frequency.

The receiver output signal is thus sampled and multiplied by a sampled version of the complex square wave $h(t)$. Such "discrete-time" multiplication is equivalent to analog multiplication followed by sampling, so we consider the latter operation sequence.
In the frequency domain, the Fourier transform of the signal is convoluted with that of the complex square wave. As a result of the convolution, the spectral line at radian frequency ω is lowered at zero frequency, and smaller-amplitude copies of it appear at $\pm 4\omega$, $\pm 8\omega$ and so on. We will refer to such copies as to "ghost lines".
If some spectral feature were present at zero frequency, it would be shifted to $-\omega$, with "ghosts" at $+3\omega$, $+7\omega$, ... and at -5ω, -9ω, etc. Also, if a spectral line were present at $-\omega$, it would be lowered to -2ω, with "ghosts" at $+2\omega$, $\pm 6\omega$, $\pm 10\omega$ and so on.
Aliasing of one of the "ghost" lines into zero frequency, and hence signal distortion, can take place if the original signal has any spectral features at frequencies $\omega \pm 4n\omega$, i.e., at 5, 9... or -3, -7... times the square wave frequency. Special care must be taken to evaluate this possibility case by case. A digital anti-aliasing filter (e.g., a lowpass with cutoff ω) could be used to prevent this problem, but such a solution can imply more computational effort than required by plain multiplication by sine waves.
We now imagine that the signal is sampled after multiplication: as a consequence, a frequency-domain "folding" takes place. Spectral features that lie outside the Nyquist band are folded, and appear as aliases in the Nyquist band.

The effect can be visualized as the following: each spectral feature is moved in constant steps until it falls within the observed band, where it adds up to the complex spectrum. If the sampling rate is t_s, the step is equal to the (radian) sampling frequency, $\omega_s = \frac{2\pi}{t_s}$.

In the situation we are considering, the "ghost" lines generated by multiplication are folded back into the observed band, and can give rise to aliasing.
The following cases can occur:

- if ω_s is a multiple of 4ω, each aliased "ghost" line is is coincident with another copy of itself, because such copies were separated by 4ω. The net effect is that the shape of each spectral line is undistorted: furthermore, the total effect on each line can be calculated because the amplitude of each "ghost line" is known (proportional to $\frac{1}{4n+1}$, as shown before;)
- if ω_s is a multiple of ω but not of 4ω, "ghosts" of the spectral lines which in the original signal corresponded at frequencies zero and -ω can be aliased into zero frequency, and distort the observed signal;
- if ω_s is not a multiple of ω, aliased "ghost" lines can spread over all the frequency field of interest, and due to the slow decay of the $\frac{1}{4n+1}$ term, strong distortion can take place.

The conclusion is that sampling frequency if ω_s should be a multiple of 4ω for minimum distortion.

We note that an even better situation occur when ω_s is exactly 4ω. In this case, the original algorithm can be implemented with small effort, because the sine wave samples reduce to:
$\sin n\omega t_s = 0, 1, 0, -1, 0, 1, 0 \ldots$
$\cos n\omega t_s = 1, 0, -1, 0, 1, 0, -1 \ldots$
so that multiplications by +1 and -1 are sufficient. This, however, is a particular case.

After sign reversal, signal is coherently integrated to select the spectral line near zero frequency. In general, a frequency range much narrower than ω will be observed.
As explained before, coherent integration can be represented as a boxcar filtering followed by decimation. The effective sampling frequency after coherent integration is $fe = \frac{fs}{N}$. Because an integer number of samples is summed, such a frequency is always a submultiple of the sampling frequency.
The signal spectrum before coherent integration contains a number of "ghosts" of the zero-frequency spectral line at $\pm 4\omega$. Further sets of ghost lines are present at $(-1 \pm 4n)\omega$ and $\pm 2n\omega$ if s(t) had features at zero frequency and at -ω.
All such lines can cause aliasing, and further distortions, because of the decimation implied in the coherent integration process.
There is no way to eliminate these lines completely by coherent integration. The problem can be reduced by considering that the first step in coherent integration is a boxcar filter with

transmission zeroes at frequencies $n\frac{fs}{N}$, corresponding to radian frequencies $n\frac{\omega_s}{N}$. If ω is a multiple of $\frac{\omega_s}{N}$, all spectral lines except the one at zero frequency fall on transmission zeroes.
The conclusion is that, for minimum distortion, the coherent integration period should be made a multiple of the square wave period.

All the above discussion holds for spectral lines. White noise is also present in virtually every radar signal, and it is important to evaluate the effects of the proposed algorithm on it. This is very easily done by considering that white noise is uncorrelated from sample to sample. Also, it can have positive or negative values with equal probability. Hence, sign reversal has no effect on noise correlation, because it produces another uncorrelated time series with the same statistics. The effects on noise of the whole process are thus the same of a plain coherent integration, namely, variance is increased N times. As explained, output noise power is multiplied by N, while signal power increases as N^2.

With careful evaluation of distorsion and aliasing effects, this simplified procedure makes double demodulation affordable even for real-time processing systems. Some problems may arise to fulfil the constraints we introduced for minimum distortion, i.e. that the sampling period, square wave period and coherent integration period be multiples of each other.
Such problems can usually be circumvented in existing radar systems. In fact:

- the sampling period is equal to the radar IPP (inter-pulse period, i.e., the time interval between succesive transmitted pulses); it is usually possible to increase the IPP, at the cost of a reduction of the mean transmitted power, to adjust it to a submultiple of the square wave period; on the contrary, it is not usually possible to reduce the IPP, because multiple-time-around echoes can appear;
- on the other hand, there is no need to use a square wave frequency strictly equal to the spectral line frequency: there can be a small difference, because the algorithm allows observation of a small frequency band;
- spectral lines are usually narrow, and rather large values of N can be used: so, it is not difficult to find a coherent integration number multiple of the required factors.

A typical parameter selection procedure can thus be the following:
- determine the spectral line frequency to observe;
- determine a square wave frequency F_q close to the spectral line; if necessary, adjust the IPP so that sampling frequency is a multiple of $4F_q$;
- determine the coherent integration period as the largest multiple of the square wave period that allows observation of the required bandwidth.

5. Possible implementations.

A possible hardware implementation of double demodulation is shown in fig.4. We consider that an integer multiple of the sampling period is used for the square wave. We further suppose that the phase of the square waves and the sample pulses are synchronized so that no sample falls on the square wave transition front. The easiest way to obtain this is to choose a square wave period multiple of 4 times the sample rate: this way, the 90 degrees phase shift between "sin" and "cos" square waves contains an integer number of samples. Thus, each square wave period is divided into four "quadrants" where the two control signals are (1,1), (-1,1), (-1,-1), (1,-1), respectively. These control signal can be generated by a two bit register, cycling through four states, clocked by a simple programmable counter. If a square wave period contains $4N$ samples, the programmable counter generates a pulse every N samples, which triggers a transition in the register state. The register bits in turn control two programmable sign switchers.

If nonmultiple periods are used, control pulses have to be generated by a separate oscillator, possibly linked to the system clock.

Another possible implementation, which avoids sign switchers, is to use the control signals as address bits to select the accumulator where samples will be added. So, all samples corresponding to a +1 control signal are accumulated in one register, and all samples corresponding to -1 in another one. At the end of the coherent integration period, the content of the second register is subtracted from the first: the result is the same as obtained by a sum with sign switching. This solution avoids sign switching but requires the use of eight accumulators: one accumulator is needed for each state (+1/-1) of each of demodulating signal ("sin" and "cos") for each of the components (real and imaginary) of the input signal.

A software implementation is also possible, provided a fast DSP is used. The program would simply generate the averaged values by accumulating N samples and subtracting the following N. If the DSP is fast enough, real multiplication by sine waves may be implemented.

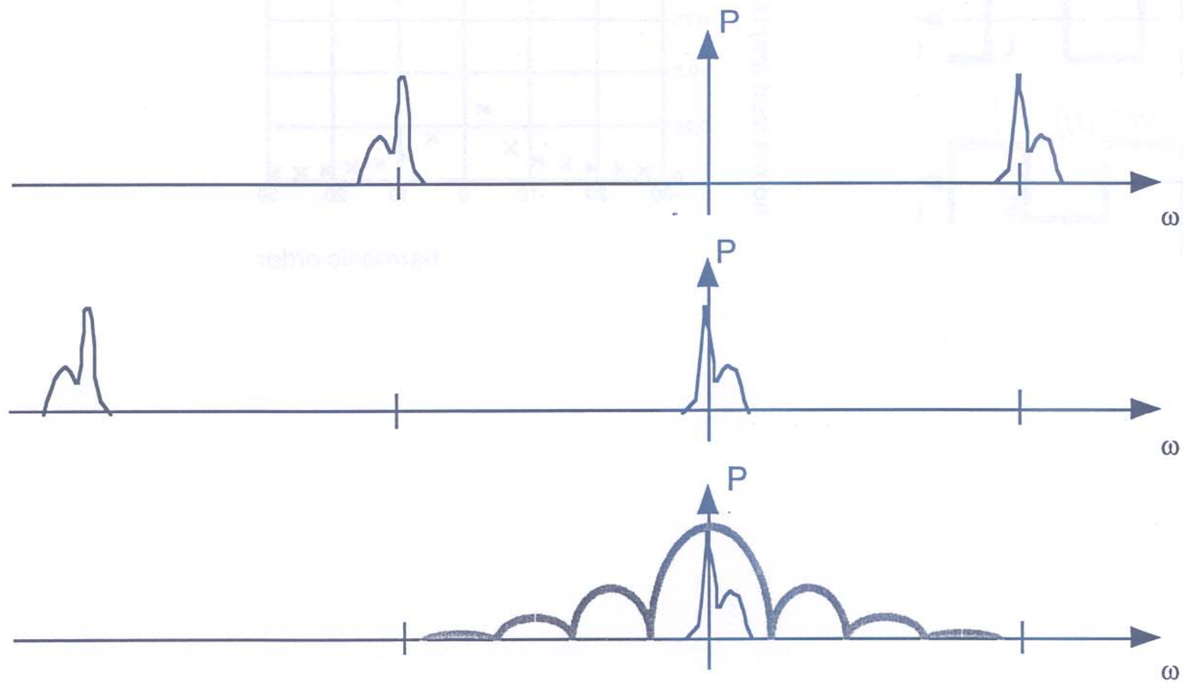
A DSP solution may be facilitated by the fact that, as explained above, four independent accumulators are used for the four results of multiplications (or sign reversal). A pipelined architecture could efficiently perform the four accumulations, because each accumulation result will only be needed after 3 other operations. As explained before, accumulator contents would then be added together after the specified number of integrations has been performed.

6. Conclusion

The double demodulation technique we propose is a digital reproposition of the analog operations performed by radar receivers to obtain baseband signals. Such an algorithm will probably find application in Doppler experiments where narrow spectral lines are measured, like radio-acoustic sounding or sea scatter experiments.

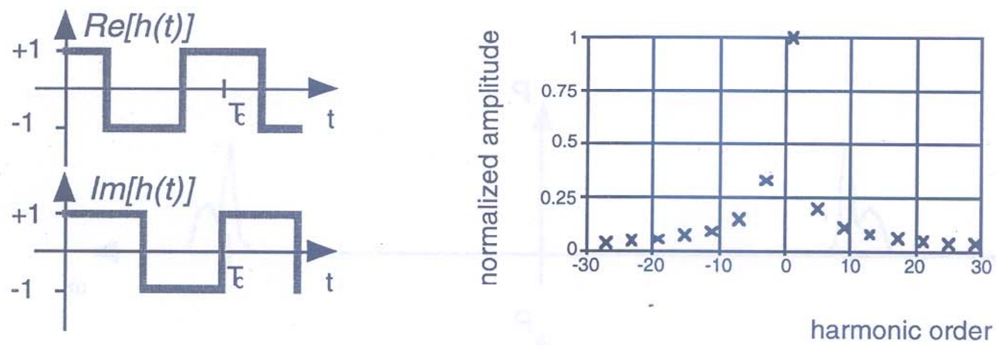
The main advantage of double demodulation consists in a reduction of computational load when high resolution frequency spectra have to be estimated. Also, it would be easy to implement such a technique, by hardware or software means, with slight modifications to existing Doppler signal processing systems.

FIGURE 1



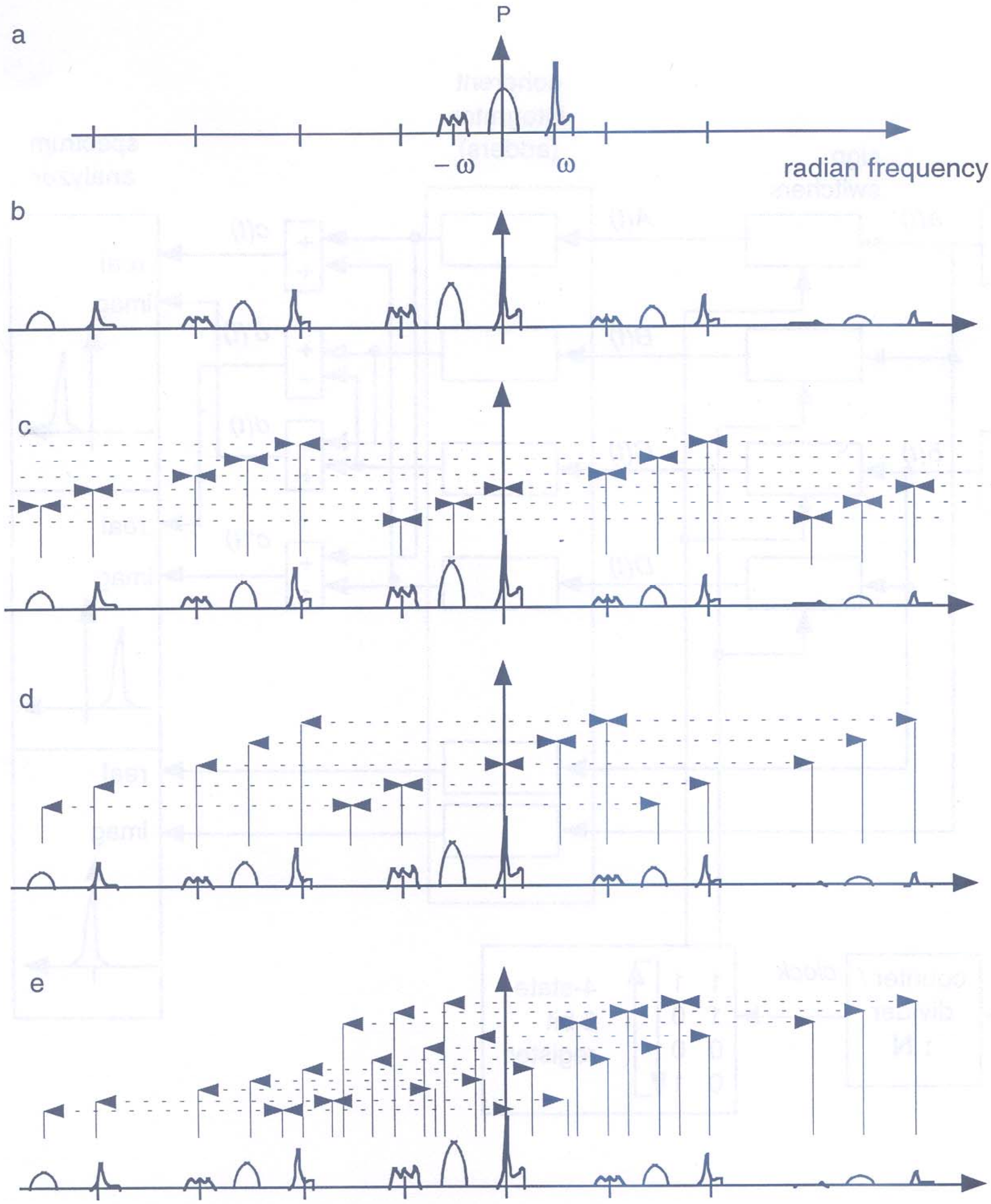
Quadrature demodulation consists of two steps: the signal is first multiplied by a demodulating factor which shifts it spectrum down in frequency, then a lowpass filter is employed to select the frequency band of interest. The resulting signal is complex, because the spectrum is not symmetric.
Double demodulation consists of the same operations, but they are performed in a digital way on a time-discrete, complex signal.

FIGURE 2



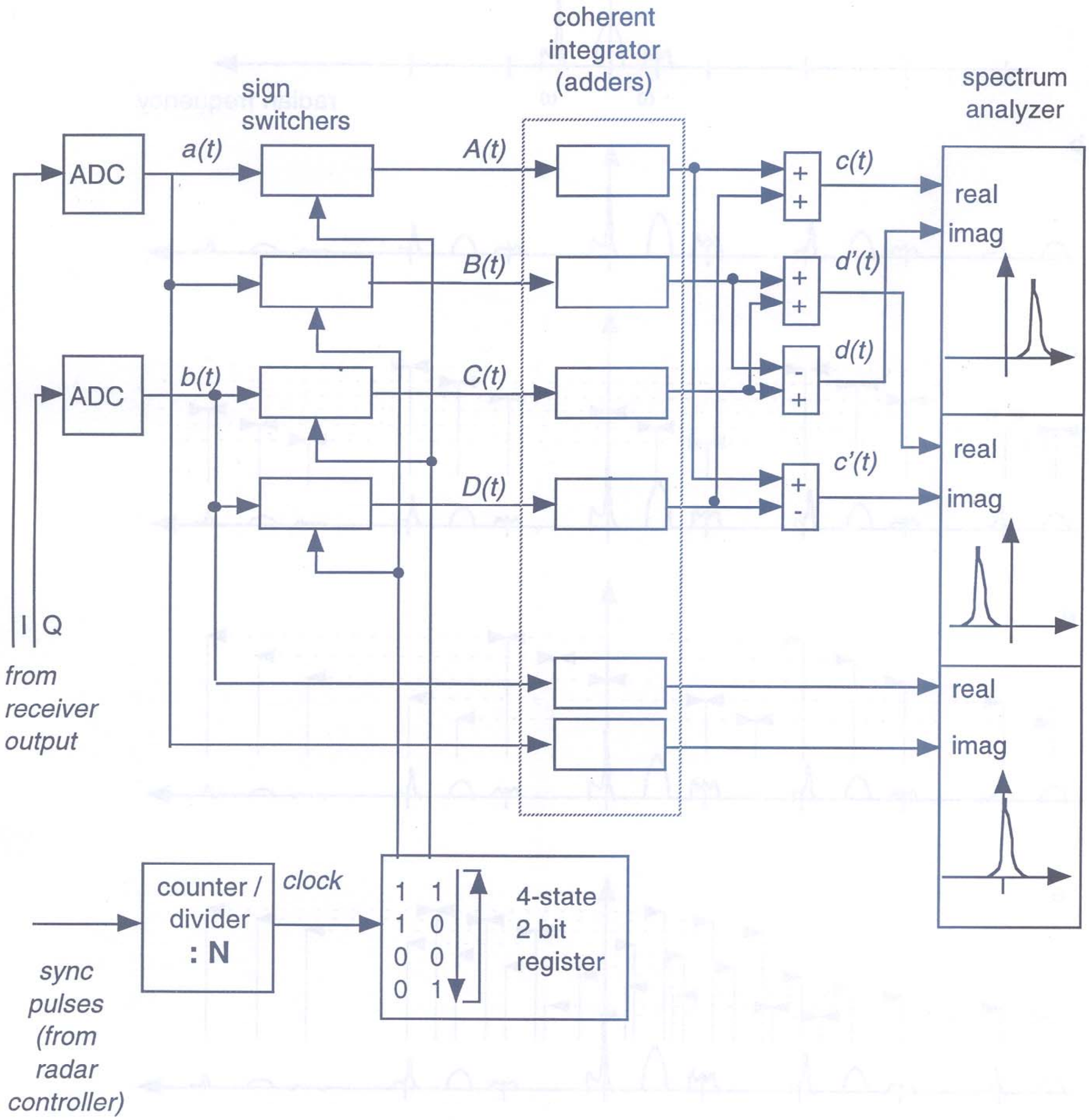
In the simplified algorithm, a "complex square wave" is used as the demodulating factor. The spectrum of a "complex square wave" has nonzero amplitude at frequencies $(1+4n)\omega$, where ω is the fundamental frequency.

FIGURE 3



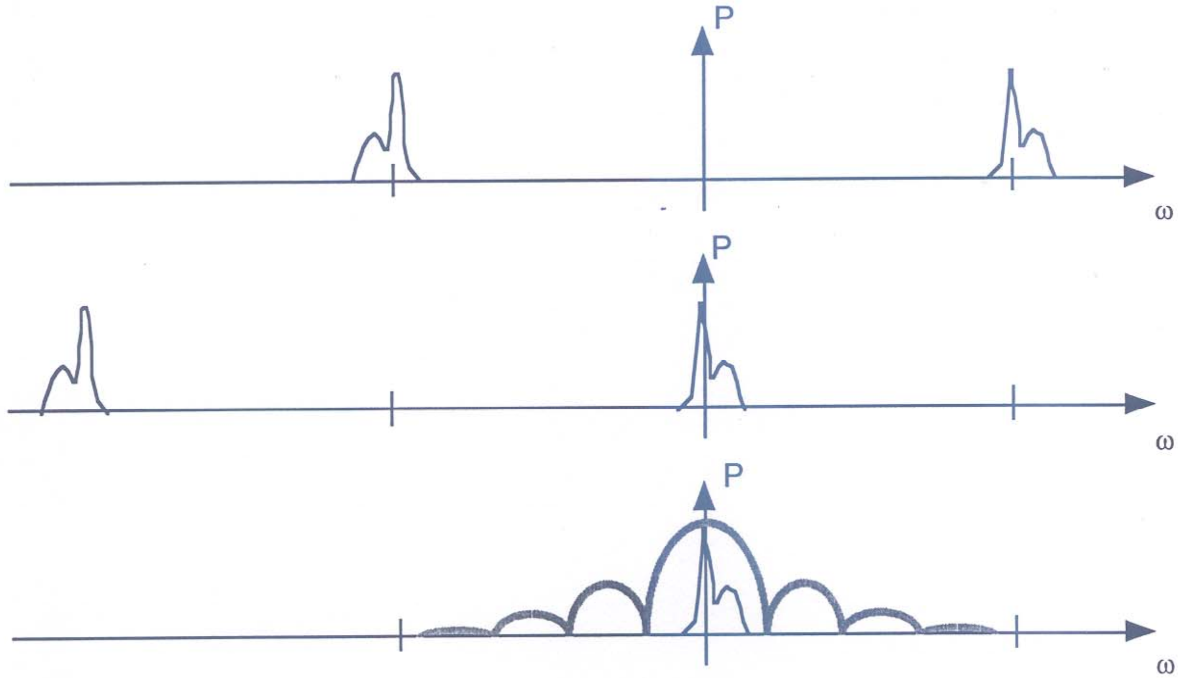
The effects of time-discrete multiplication by a complex square wave can be visualized considering a continuous-time multiplication followed by sampling.
Let us consider a signal like in (a), which has spectral features at frequencies ω and -ω, as well as at zero frequency.
Continuous-time multiplication by a square wave of frequency ω causes an infinite set of "ghost lines" to appear (b).
Sampling can have different effects: if the sampling frequency is an integer multiple of 4 times the square wave frequency, each line is coincident with another copy of itself, and no net aliasing takes place (c); if the sampling frequency is a multiple of ω but not of 4ω, or if it is not multiple of ω, the spectrum is distorted due to aliasing ((d) and (e)).

FIGURE 4



A possible hardware implementation of double demodulation using complex square waves

FIGURE 1



Quadrature demodulation consists of two steps: the signal is first multiplied by a demodulating factor which shifts it spectrum down in frequency, then a lowpass filter is employed to select the frequency band of interest. The resulting signal is complex, because the spectrum is not symmetric.
Double demodulation consists of the same operations, but they are performed in a digital way on a time-discrete, complex signal.