IEE TELECOMMUNICATIONS SERIES 5

Angle Modulation

IEE TELECOMMUNICATIONS SERIES 5

SERIES EDITOR: PROF. J.E. FLOOD

Angle modulation:

the theory of system assessment

Published by Peter Peregrinus Ltd.,
Southgate House, Stevenage, Herts. SG1 1HQ, England

ISBN: 0 901223 95 6

Typeset by the Alden Press (London and Northampton) Ltd.,
Printed in England by A. Wheaton & Co., Exeter

Previous volumes in this series

Telecommunication networks
J.E. Flood (Editor)

Principles of telecommunication-traffic engineering
D. Bear

Programming electronic switching systems
M.T. Hills and S. Kano

Digital transmission systems
P. Bylanski and D.G.W. Ingram

Angle modulation:
the theory of system assessment

J.H. Roberts, B.Sc., F.I.M.A.
Principal Mathematician
Plessey Avionics & Communications
Roke Manor
Romsey,
Hants., England

PETER PEREGRINUS LTD.
on behalf of the
Institution of Electrical Engineers

Preface

Angle modulation is the generic term applied to phase modulated (p.m.) and frequency modulated (f.m.) transmissions, and a large variety of present day communication systems use this means of conveying information in analogue or digital form, and, sometimes, in both.

The initials f.m. and p.m. are often encountered in the fields of Radar, T.V. local and police radio, satellite transmissions etc., and the rival amplitude modulation (a.m.) is now much less used. The early history of f.m. with its fervent championing in the USA by Major Armstrong (sometimes hampered, apparently, by equipment manufacturers who saw their sales slumping if the freedom from interference of f.m. compared with a.m. became too well known) and its misunderstandings regarding bandwidth, which were finally resolved by Carson, makes interesting reading.[1-3]

With both f.m. and p.m. the processes of modulation and demodulation are nonlinear so the analysis of performance of f.m. and p.m. systems when disturbed by noise and interference has long presented an attractive challenge to the analyst. This inherent nonlinearity means that, when decision theory is used to identify the optimum method of combatting noise and various types of system disturbances, a complicated receiver structure can emerge that is both difficult to analyse and build, but it would appear that, in many cases, conventional designs are near-optimum. Usually, rather special reasons have to exist for the optimum detector to be implemented and, more often, constraints imposed by date-lines and budgets force the designer to opt for tried and tested methods of detection, with the occasional variation in a conventional design. The analysis to be given here refers mainly to this latter situation.

Another consequence of the nonlinear nature is that the theory has developed in a piecemeal fashion with different approaches being tried from time to time. When a result is reported that accounts exactly for the nonlinearity, then this is particularly satisfying, but often a tractable theory can only be produced by taking steps that a determined critic can object to, and justification is strengthened by the results of computer simulation runs or laboratory measurements that are done to check the theoretical predictions. This is particularly the case when interest lies in the outcome of first filtering an f.m. wave and then demodulating it, and a survey of this problem is given here.

The monograph is principally concerned with the harmful effects of noise and interference, and many useful results from the large literature that now exists are quoted. It has become customary to include an extensive bibliography of pertinent literature, but it is becoming increasingly difficult to keep up to date (one recently published survey listed over eight hundred references on the phase lock loop alone) and here a unifying theory is developed which gives formal solutions to the performance assessment problems that are posed and which allows previously derived results to be credited to their originators at appropriate points during an overview of the subject.

The reader is assumed to be a practising engineer or graduate student and therefore familiar with such communication engineering concepts as power spectra, coherent and incoherent detection, Gaussian noise, error rates etc., but some of the mathematical tools used, such as the signum function ($\operatorname{sgn} x = \pm 1$ according as x is positive or negative) or the Dirac delta function $[d \operatorname{sgn} x/dx = 2\delta(x)]$ may be a little unfamiliar. The Dirac delta is the most commonly encountered generalised function, and the discipline of generalised function theory, as described by Lighthill, is tacitly assumed here to gain freedom regarding multiple operations such as interchanging orders of integration or differentiating under the integral sign, and such steps will be taken without further comment with all limits being assumed to exist. A willingness to grapple with the double and quadruple integrals that need to be evaluated is therefore desirable but the aim is always to present a final result in as neat a form as possible so that its physical significance is apparent and it is usable to the reader. Moreover, the general availability of small but powerful desk calculators has greatly widened the class of performance formulae that can be easily handled.

In recognition of the increasing amount of data that is now transmitted in digital form, the subject matter is divided about equally between analogue and digital f.m. and p.m. signals, and the first three chapters develop the underlying theory. Then attention is given to predicting the performance of particular transmissions when disturbed by noise or adjacent channel interference in the situations commonly encountered by the systems analyst.

References

1 TUCKER, D.G.: 'The invention of frequency modulation in 1902', *J. Inst. Electron. & Radio Eng.*, 1970, **40**, (1), pp. 33–37
2 TUCKER, D.G.: 'The early history of amplitude modulation, side bands and frequency-division-multiplex', *J. Inst. Electron. & Radio Eng.*, 1971, **41**, (1), pp. 43–47
3 ANGUS, R.: 'Major Armstrong and the struggle for FM broadcasting', *Audio Scene Can.*, June 1974, pp. 30–34

Acknowledgments

The invitation to write this volume came from the IEE, and the author thanks them, and Dr. R. Hamer in particular, for the opportunity it afforded.

It is a pleasure to thank my colleagues for their help, David Randall for discussions and valuable checking, and especially Frank Gardner and Keith Wales who undertook to read the drafts and subsequently made a number of improvements.

At several points reference is made to computer programs but little attention is given to the detailed and careful work that lies behind them. For these I am indebted to Ann Griffiths.

The drawings were prepared by Neil Martin, and the general facilities made available to me by the Plessey Company through the offices of S.M. Cobb, are gratefully acknowledged.

Finally, I thank my wife Diana who typed the bulk of the manuscript and who never failed me in her support and encouragement.

Chandler's Ford September 1976

List of principal symbols

Unless otherwise stated, frequencies are expressed in rad/s, and all power spectra as power/rad/s.

A	Carrier amplitude
a_1, a_2	Parameters of Tikhonov distribution
a_n, b_n	Coefficients in power series expansions of transfer characteristic
$\alpha(t)$	Phase modulation on unwanted transmission
$2\alpha p_m$	3 dB bandwidth of single-pole bandpass filter
βp_m	3 dB bandwidth of single-pole lowpass filter in d.t.f.
$B(t)$	Envelope of unwanted transmission
$B_1(\omega), \lambda$	Coherence factors (voltage ratios)
$2\omega_B$	Carson bandwidth
$c(t)$	Control voltage in d.t.f.
$\delta(x)$	Dirac delta function
w_d	r.m.s. test-tone deviation
ω_D	Carrier frequency separation
D/S	distortion-to-signal (power ratio)

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\, du$$

F	Feedback factor (voltage-wise)
$g(t)$	Impulse response function of equivalent lowpass filter
$\left\{\dfrac{G(q)}{\frac{1}{2}A^2}\right\}$	Normalised r.f. power spectrum
h	Time interval that is made vanishingly small
$H(j\omega)$	Transfer characteristic of equivalent lowpass filter
$He_n(x)$	Hermite polynomial
I_c, I_s	In-phase and quadrature components of narrow-band Gaussian noise (understood to be time dependent)
$I_e(k, x)$	Rice's I_e function (see Appendix)

$I_n(x)$	Modified Bessel function of the first kind, of order n
j	$\sqrt{-1}$
$J_n(x)$	Bessel function of the first kind, of order n
k	a.m. to p.m. conversion factor (Section 5.9 and Section 8.2)
k	Binary f.m. deviation ratio (peak-to-peak frequency excursion, Hz, divided by the data rate)
L	Threshold
$2p_l$	Bandwidth of truncating filter
$L(j\omega)$	Transfer characteristic of lowpass filter
m	Parameter of Nakagami's m distribution
$M(t)$	Frequency modulation
M	Value of $M(t)$ at specific instant
$\mu(t)$	Phase modulation
$\dot{\mu}(t) = M(t)$	Frequency modulation
ν	Expected number of upward and downward f.m. clicks
n.p.r.	Noise power ratio
$\psi_N(0)$	Noise power
$\psi_N(\tau)$	Autocorrelation function of lowpass noise
p_m	Maximum modulating frequency
p_0	Minimum modulating frequency
p, q, ω	General baseband frequencies
$P(p)$	Pre-emphasis law (power-ratio) applied at baseband frequency p
$P(t), Q(t)$	In-phase and quadrature components of general narrow-band transmission
$Q(a, b)$	Marcum's Q function
$\phi(t)$	PLL phase error (Chapter 9)
$\phi(\omega)$	Phase characteristic of equivalent lowpass filter
r	Relative amplitude of interfering wave or relative echo amplitude
$r_N(\tau) = \dfrac{\psi_N(\tau)}{\psi_N(0)}$	
$R(\tau)$	Autocorrelation function
$R(t), R_s(t)$	Envelope functions
ρ	Carrier-to-noise (power ratio)
sgn (x)	$+1$ if $x > 0$, -1 if $x < 0$ zero if $x = 0$
$s(t)$	Signalling pulse in digital phase modulation
$S(\tau)_e$	Triangular wave
τ_0	Echo delay
T	General time period (Chapter 3)
T	Bit duration ($1/T$ Hz is the data rate: Chapter 10)

T_1	Interval of time that contains many carrier cycles but is not long enough for the in-phase and quadrature components to have changed significantly
$\theta(t), \theta_s(t)$	Phase functions
u	Dummy variable
$\Phi(u)$	Univariate characteristic function
u_1, u_2	Dummy variables
$\Phi(u_1, u_2)$	Bivariate characteristic function
$W_\mu(\omega)$	Power spectrum of $\mu(t)$
$\psi_\mu(\tau)$	Autocorrelation function of $\mu(t)$
$W_M(\omega)$	Power spectrum of $M(t)$
$\psi_M(\tau)$	Autocorrelation function of $M(t)$
$W(\omega)$	General power spectrum
$2\Delta_\omega$	3 dB static bandwidth of the d.t.f.

$$\Delta(h) = \theta_s\left(t + \frac{h}{2}\right) - \theta_s\left(t - \frac{h}{2}\right)$$

$$\Delta_s = \mu(T) - \mu(0)$$

ω_0	Carrier frequency
ω_I	Intermediate frequency
$\omega_i(t)$	Instantaneous frequency
ω_Δ	r.m.s. noiseband frequency deviation
$Z(j\omega)$	Loop transfer characteristic
Z_0	Expected number of zeros per second

Contents

Chapter 1

Introduction

The signal, noise, and interference waveforms to be considered are of the narrow-band type, by which it is meant that the carrier or intermediate frequency used greatly exceeds the occupied bandwidth. The general expression for a narrow-band wave, which will appear repeatedly in various guises through the monograph, is written as follows

$$V(t) = P(t)\cos\omega_0 t - Q(t)\sin\omega_0 t \qquad (1.1)$$

Here $P(t)$ and $Q(t)$ are called the in-phase and quadrature components, respectively, and it is important to appreciate two time scales (or, equivalently, two rates of change) associated with eqn. 1.1. The narrow-band assumption means that $P(t)$ and $Q(t)$ are to be regarded as slowly varying in the sense that in the time T_1, say, taken by $P(t)$ and/or $Q(t)$ to change significantly (one-half the reciprocal of the occupied bandwidth is representative of such a time period) a large number of cycles to $\cos\omega_0 t$ have gone by. The particular large number in question depends on the ratio of centre frequency (ω_0 rad/s) to the bandwidth that happens to be used, but many hundreds or many thousands would be typical; great precision is not necessary with this aspect of the theory and most of our results refer to the limiting situation in which the ratio of occupied bandwidth to centre frequency is vanishingly small. A consequence of the wide disparity in time scales is that it is often appropriate to perform an average over the time T_1 (thus smoothing out the fast variations) and produce a result which is time dependent but which varies at the slower rate. Of course, if a long-term time average is taken across many intervals of length T_1 [symbolised by $\lim_{n\to\infty}(nT_1)^{-1}\int_0^{nT_1}(\quad)dt$], then only the d.c. term, if any, is obtained.

Returning to eqn. 1.1, $[P^2(t) + Q^2(t)]^{1/2}$ is the envelope and $\tan^{-1}(Q(t)/P(t))$ is the instantaneous phase, and a representation such as eqn. 1.1 can be used for the transmitted signal as well as the wave appearing at the receiver terminals after limiting, amplification, filtering, and the addition of noise or interference have taken place.

The undistorted, unfiltered, and interference free angle modulated signal has the form

$$S(t) = A \cos [\omega_0 t + \mu(t)] \tag{1.2}$$

where $\mu(t)$ is the information-bearing phase modulation and the frequency modulation is the time derivative of $\mu(t)$, i.e. $\dot{\mu}(t)$. It will be supposed that some form of additive narrow-band interference accompanies this wave so that the receiver input is given by

$$\begin{aligned} V(t) &= A \cos [\omega_0 t + \mu(t)] + X(t) \cos \omega_0 t - Y(t) \sin \omega_0 t \\ &= R(t) \cos [\omega_0 t + \theta(t)] \end{aligned} \tag{1.3}$$

$R(t)$ is the envelope and $\theta(t)$ is the instantaneous phase. The following identifications can then be made

$$\left.\begin{aligned} R(t) &= \{[A \cos \mu(t) + X(t)]^2 + [A \sin \mu(t) + Y(t)]^2\}^{1/2} \\ \theta(t) &= \tan^{-1}\left\{\frac{A \sin \mu(t) + Y(t)}{A \cos \mu(t) + X(t)}\right\} \end{aligned}\right\} \tag{1.4}$$

An investigation of system performance may be regarded as an exercise in which certain statistical features of eqn. 1.3 are compared with the corresponding features of eqn. 1.2 and some agreed yard-stick (such as a distortion level or an error-rate) is calculated as a measure of the comparison. Depending on the demodulation method (coherent or incoherent) and the form of the information transmitted (whether $\mu(t)$ represents an analogue signal or is a pulse train representing the digitised form of some analogue signal) so the method of calculation of the yard-stick can differ and, typically, interest may centre on one or more of the following:

(*a*) the spectral density of $\dot{\theta}(t)$
(*b*) the probability distribution of $\dot{\theta}(t)$
(*c*) the probability distribution of $\theta(t + \tau) - \theta(t)$ (τ a general delay)
(*d*) the expected number of zeros of $V(t)$

Also, with the radio spectrum becoming more and more congested there is often a need to establish the band over which significant frequency components of eqn. 1.2, or perhaps the hard limited version of eqn. 1.3, i.e. $K \cos [\omega_0 t + \theta(t)]$, extend. The latter could be of interest when $X(t)$ and $Y(t)$ (see eqn. 1.3) represent sums of the in-phase and quadrature components of transmissions that share the use of a hard limiting satellite transponder with a wanted signal $A\cos[\omega_0 t + \mu(t)]$.

For the major part of our development only second-order statistical quantities have to be calculated, perhaps the most familiar of which is the autocorrelation function. For $V(t)$ this is written

$$R(\tau) = \langle V(t)V(t+\tau)\rangle \tag{1.5}$$

The angle brackets $\langle\ \rangle$ indicate an appropriate time average and an ensemble average taken over any random parameters of $V(t)$ (associated with modulation, for example). The left-hand side of eqn. 1.5 has been shown as a function of τ only as this will be true of the cases to be met here. However, this does not necessarily imply that $V(t)$ is stationary.

By invoking the Wiener-Khinchine relationship the spectral density of $V(t)$ [or power spectrum: $W(\omega)$ $\omega \geqslant 0$] is given by taking the Fourier transform of $R(\tau)$

$$W(\omega) = \frac{2}{\pi}\int_0^\infty R(\tau)\cos\omega\tau d\tau \tag{1.6}$$

With present day computing facilities this integral seldom presents serious difficulty and here attention will be focussed on calculating $R(\tau)$ in particular cases of interest.

The brackets $\langle\ \rangle$ will appear on numerous occasions and are not confined to use when finding autocorrelation functions. Their meaning will be clear from the context, and in some cases the time averaging operation is not required. An example of this arises when item (*b*) (the distribution of $\dot\theta(t)$) is under investigation.

If a signal, such as $\dot\theta(t)$, is sampled at some time instant $t = t_0$ then $\dot\theta(t_0)$ will often have a range of possible values. The probability $Pr[\dot\theta > L]$ that $\dot\theta(t_0)$ exceeds a threshold L can then be expressed as

$$Pr[\dot\theta > L] = \langle \tfrac{1}{2}\{1 - \mathrm{sgn}[L - \dot\theta(t_0)]\}\rangle \tag{1.7}$$

Here $\mathrm{sgn}(X) = \pm 1$ according as X is positive or negative and the brackets $\langle\ \rangle$ evidently refer to an averaging operation in which each possible value of $\dot\theta(t_0)$ is inserted in the expression $\tfrac{1}{2}\{1 - \mathrm{sgn}[L - \dot\theta(t_0)]\}$ and the resulting 0 or 1 is weighted by the probability of that particular $\dot\theta(t_0)$ arising. An alternative way of writing eqn. 1.7 is

$$Pr[\dot\theta > L] = \int_{-\infty}^{+\infty} \tfrac{1}{2}\{1 - \mathrm{sgn}[L - \dot\theta(t_0)]\}\cdot p[\dot\theta(t_0)]\, d[\dot\theta(t_0)] \tag{1.8}$$