

THE UNIVERSITY OF LEEDS
School of Electronic and Electrical Engineering

**Channel Characterisation
and System Design for
Sub-Surface Communications**

by
Anthony David Wilson Gibson, M.A.

February 2003

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

At Leeds University, I am, of course, extremely grateful to my supervisor **Prof. Mike Darnell** for his help and guidance; and also to **Dr Hal Strangeways**, with whom I have also had some stimulating debates and to **Dr Chris Trayner** and **Dr Paul Clark** for their advice and guidance.

David Brenkley of the Mines Rescue Service, and **David Lewis** of International Mining Consultants Ltd, for both of whom I have acted as a consultant, have given me some invaluable background on mine communications and have encouraged me to further pursue this work beyond what is reported in this thesis.

The British Cave Research Association's *Cave Radio & Electronics Group* and its USA-based 'sister' group *Speleonics* (a division of the USA National Speleological Association) both contain a mix of cavers, radio amateurs, electronics hobbyists, and professional engineers and scientists, many of whom have assisted me by means of correspondence, debate at field meetings and discussion on the Internet in various Usenet and mailing-list forums. In particular there has been much useful debate with my CREG colleagues **Mike Bedford**, **John Hey**, **Derek Potter**, **Chris Trayner**, **John Rabson**, and **Richard Rushton**.

I have also been in correspondence with cave radio experimenters **Ian Drummond** from Canada and **Brian Pease** from the USA; **Steven Shope** supplied me with a copy of his field-evaluation program; **Dr John Frankland** outlined CRO medical practices for me; **Peter Dodd** (editor of Radcom magazine), **Joe Hussey** and **Prof. Roger Jennison** gave me information on anapole (toroidal) radiating antennas.

CREG's publication, *The Bibliography of Underground Communications* by **Chris Trayner** and **Nick Williams**, has been a useful source of reference. **Chris Trayner** loaned me a number of items from the CREG library. **Joe Giddens** supplied me with a copy of Watt's book *VLF Radio Engineering*. **Prof. James Wait** (!) gave me a signed copy of his book, *Geo-Electromagnetism*. Others – in particular **Chris Howes** (editor of Descent), **Ray Mansfield** (editor of Speleo Abstracts), **Nick Negus**, **Dr Mark Noel** and **Guy Tomlinson** – assisted by passing information to me. If I forgotten anyone – I apologise.

In the absence of any sponsorship or grant, the costs of these studies were met entirely from personal funds, and I am therefore extremely grateful to my partner, **Ann Whilding**, for supporting me over this lengthy period.

This thesis was prepared using **Microsoft Word 97/SR2**, and assisted by software applications **Endnote 5.0**, **CorelDraw 5.0**, **MatLab Release 12 (Student Version)** **Paint Shop Pro 4.12 shareware edition**, **Microsoft Excel 97/SR2** and **Nettlist** from Racal-Redac Systems.

Hofstadter's Law: It always takes longer than you expect,
even when you take into account Hofstadter's Law.

Abstract

Channel Characterisation and System Design for Sub-Surface Communications

Anthony David Wilson GIBSON

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy, February 2003

Sub-surface or through-the-earth communication using electromagnetic fields – and specifically magnetic induction equipment – plays a key role in search and rescue systems used in the mining industry and, increasingly, by cavers and pot-holers. Similar equipment is used for radio-location, sub-surface surveying and geophysical measurements.

The use of fast desktop computers allows a mathematical model of the propagation to be investigated in detail, demonstrating the preferred orientation of the antennas and the existence of an optimum frequency that is dependent on depth and other parameters. Computer simulations demonstrate a reduction in accuracy of radiolocation at skin depth distances, and methods of correcting this inaccuracy are introduced.

The transmitter and receiver antennas for portable induction loop systems are usually air-cored loops or magnetic-cored solenoids, tuned to resonance. However, the preferred antenna often depends on the intended use of the communication system, with untuned antennas having an advantage in some situations. The use of toroid (anapole) structures and rotating magnets as transmitters is discussed briefly. A figure of merit – the specific aperture – is introduced as an aid to antenna design.

Internal (amplifier) noise can be reduced by noise-matching, although special account has to be taken of the inductive antenna. Internal noise is often swamped by external (atmospheric) noise, for which several mitigation strategies are discussed. Differing up-link and down-link noise performance may dictate different antennas. External noise is frequently characterised by the atmospheric noise temperature ratio, but data derived using the standard electric field antenna cannot be applied to a study of magnetic noise.

The design of a wide-band low-frequency channel sounder is described, with which it is intended to perform a detailed channel evaluation using a binary sounding sequence. A simple method of calculating the inverse of such a sequence is introduced, for which cross-correlation with the inverse sequence at the receiver results in a system identification signal that is used to maintain synchronism with the transmitter. The extreme wideband nature of the system results in a low efficiency, which is countered by using signal-averaging techniques at the receiver. Preliminary results are reported, in which the sounder was used to capture background noise.

Keywords: atmospheric noise, cave radio, channel sounding, electromagnetic theory, inverse sequence, mine rescue, noise-matching, noise temperature, propagation, radiolocation, sequence design, specific aperture, sub-surface communication.

Contents

Acknowledgements	ii
Abstract	iii
Contents	iv
Glossary of Symbols and Abbreviations	xiii
List of Figures	xix
List of Tables	xxi

1 SUB-SURFACE COMMUNICATIONS		1
1.1	Uses for Sub-Surface Communications	1
1.1.1	Mine Applications	1
1.1.2	Cave and Geophysical Applications	2
1.2	Features of Sub-surface Radio	2
1.2.1	Mine Communications	3
1.2.2	Cave Communications	3
1.3	Aims of this Study	5
1.4	Sub-Surface Penetration of Electromagnetic Fields	6
1.4.1	Propagation	6
1.4.2	Antennas	7
1.4.3	Near and Far Fields; Transition Distance	7
1.4.3.1	The Effect of a Conducting Medium	9
1.4.4	Modelling the Field Equations	10
1.4.4.1	Fields in a Uniform Infinite Conducting Medium	10
1.4.4.2	Fields in a Uniform Conducting Half-Plane	10
1.4.5	Bibliography	11
1.4.6	Is Sub-Surface 'Radio' a Radio Phenomenon?	13
1.5	Example of a Communications Link	15
1.5.1	Transmitter Antenna Design	15
1.5.2	Receiver Antenna	16
1.5.3	Propagation Losses	17
1.5.3.1	Up-link operation	17
1.5.3.2	Down-link operation	18
1.5.4	Optimum Frequency	18
1.6	Chapter Summary and Original Contributions	19
1.7	References	22
2 PROPAGATION		24
2.1	Electric v. Magnetic Fields	24
2.2	A Simple Model of a Loop Antenna	25
2.2.1.1	Far or Radiation Field	27
2.2.1.2	Near or Induction Field	27
2.2.1.3	Observations on the Inverse Cube Law	28
2.2.2	Propagation in a Conducting Medium	29
2.2.2.1	Plane Wave Propagation and Skin Depth	29

2.2.2.2	Electrical and Magnetic Properties of the Ground	31
2.2.3	Fields from a Magnetic Dipole in an Infinite Conducting Medium	33
2.2.3.1	Validity of Near Field Approximation	36
2.2.3.2	Definition of Quasi-Static and other Field Approximations	37
2.2.4	Derivation of an Optimum Frequency	39
2.3	A Model for a Uniform Half-Plane	41
2.3.1.1	Boundary Conditions and Computer Simulation	42
2.3.2	Orientation of Transmitter and Receiver	43
2.3.2.1	Classification of Antenna Alignments	44
2.3.2.2	Classification of Communications Paths	46
2.3.3	Field Expressions	47
2.3.3.1	Horizontal Loops (VMD) in coaxial alignment	47
2.3.3.2	Modification for Down-Link Fields	48
2.3.3.3	Vertical Loops (HMD) in Co-planar Alignment	48
2.3.3.4	Elevated Source and Field Point	48
2.3.3.5	Orthogonal Loops – Study of Secondary Field	49
2.3.4	Computer Simulation of a Buried VMD	49
2.3.4.1	Comparison of half-space and full-space models	50
2.3.4.2	Q-field v. offset D for a Range of Ground Conductivity, \mathfrak{F}	50
2.3.4.3	P-field v. offset D for a Range of Ground Conductivity, \mathfrak{F}	52
2.3.5	Optimum Frequency using the Half-Plane Model	53
2.4	Other Aspects of Propagation	55
2.4.1	Choice of Antenna System	55
2.4.2	Dispersion	55
2.4.3	Multipath Propagation	55
2.4.4	Geophysical Methods	56
2.4.5	Radiolocation	56
2.4.5.1	Radiolocation at Skin-depth Distances	57
2.4.5.2	The 'Tilted Body' Problem	57
2.5	Concluding Remarks	58
2.6	References	59
3	SPECIFIC APERTURE:	
	A FIGURE OF MERIT FOR INDUCTION LOOP ANTENNAS	61
3.1	Deriving a 'Figure of Merit'	61
3.1.1	Demonstration of Principle	62
3.1.2	Specific Aperture	62
3.1.2.1	Prior Art	64
3.2	Observations using 'Specific Aperture'	64
3.2.1	Typical Values	64
3.2.2	Field Strength is not Solely Related to Power	65
3.2.3	Number of Turns has no Effect on Performance	65
3.2.4	Shape of Antenna	66
3.2.5	Aluminium is Better than Copper	67
3.2.5.1	Application to E-field Antennas	67
3.2.6	Q-Factor does not Depend on Number of Turns	68
3.3	Using 'Specific Aperture' with a Receiving Antenna	69
3.3.1	Thermal Noise	70
3.3.2	Other Sources of Noise	70
3.4	Using 'Specific Aperture' with a Radiating Antenna	71

3.4.1	Radiation Resistance	71
3.4.1.1	Application to SNR equation	71
3.4.1.2	Efficiency of Radiating Antennas	72
3.4.1.3	Radiation Resistance v. Number of Turns	72
3.4.2	Effective Aperture and Effective Height	73
3.4.3	Buried Transmitter Loops	74
3.5	Concluding Remarks	74
3.6	References	75
4	ANTENNA DESIGN	76
4.1	Antenna Classification	76
4.1.1	H-field Antennas	77
4.1.2	E-field Antennas	78
4.1.2.1	Comparison of E- and H-field Antennas	79
4.1.3	J-field Antennas	82
4.1.3.1	The HeyPhone	84
4.1.4	Cable-based Communications	85
4.1.4.1	Two wire telephones	85
4.1.4.2	Single-wire telephones	85
4.1.4.3	Hybrid with Induction Radio	85
4.1.4.4	Guide-Wire Radio & Leaky Feeders	85
4.2	Choice of Antenna System	86
4.3	Designing Induction Loops using Specific Aperture	86
4.3.1	Application to Ferrite-cored Antennas	87
4.3.2	Further Observations using Specific Aperture	88
4.3.2.1	Sixth-power Law for Range	88
4.3.2.2	Third-power Law for Range	89
4.3.2.3	Air-cored Antenna Arrays	89
4.3.2.4	Matching the Antenna Mass	90
4.4	Magnetic-cored Antennas	91
4.4.1	Magnetic Moment (Transmitter)	92
4.4.2	Demagnetising Factor	93
4.4.2.1	General Expression	93
4.4.2.2	Approximation for Cylindrical Solenoid	94
4.4.2.3	Ferrite Tubes	95
4.4.2.4	Optimum Shape Factor	95
4.4.3	Specific Aperture and Inductance	95
4.4.4	Power Dissipation (Transmitter)	96
4.4.4.1	Winding and Core Losses	96
4.4.4.2	Linearity	96
4.4.5	Antenna Arrays	97
4.5	Air-core v. Magnetic-cored Antennas	99
4.5.1.1	Air-cored Loop	99
4.5.1.2	Ferrite-cored Solenoid	99
4.6	Coupling to the Ground	102
4.7	Proximity and Skin Effects	102
4.8	Exotic Antennas	103
4.8.1	Toroidal Antennas (Transmitter)	104
4.8.2	Rotating Permanent Magnet (Transmitter)	106
4.8.2.1	Magnetic Moment	106

4.8.2.2 Demagnetising Factor	106
4.8.2.3 Screening	107
4.8.2.4 Radiation and External Resistance	107
4.8.3 Specialised Magnetic Materials	108
4.9 Concluding Remarks	109
4.10 References	110
5 TRANSMITTER DESIGN	112
5.1 Antenna Comparisons	112
5.1.1 Tuned v. Untuned Antennas	112
5.1.2 Air-core v. Ferrite-core Antennas	113
5.1.3 Comparison by Power Dissipation	113
5.2 Driving a Tuned Antenna	113
5.2.1 Matching the Antenna to the Amplifier	113
5.2.2 A Figure of Merit for Bandwidth Comparisons	115
5.2.2.1 Equalisation	116
5.2.2.2 Tuning Capacitor	116
5.2.2.3 Frequency Drift	118
5.3 Driving an Untuned Antenna	118
5.3.1 Analogue Drivers	118
5.3.1.1 Driving a Partially-tuned Antenna	119
5.3.2 Digital Drivers	119
5.3.3 A Figure of Merit for Untuned Antennas	120
5.3.3.1 Analogue Drivers	121
5.3.3.2 Digital Drivers	121
5.3.4 Comparisons & Remarks	122
5.3.4.1 Analogue Signals	122
5.3.4.2 Digital Signals	122
5.4 Damped Antennas	122
5.4.1 Damping a Tuned Antenna	123
5.4.2 Damping an Untuned Antenna	123
5.4.2.1 Analogue Driver	123
5.4.2.2 Digital Driver	124
5.5 Modulation Methods	124
5.5.1 Analogue Methods	124
5.5.1.1 Single Sideband	124
5.5.1.2 Carrier Methods (AM, FM)	124
5.5.1.3 Double Sideband, Suppressed Carrier	125
5.5.2 Digital Methods	125
5.6 Concluding Remarks	125
5.7 References	127
6 NOISE AND RECEIVER DESIGN	128
6.1 Fundamentals of Antenna Design	129
6.1.1 Antenna Comparisons	129
6.1.1.1 Tuned v. Untuned Antennas	129
6.1.1.2 Air-core v. Ferrite-core Antennas	129
6.1.2 Signal / Noise Ratio	129
6.1.2.1 A General Expression	129
6.1.2.2 Noise Factor	131

6.1.2.3 Designing using the Noise Factor	132
6.2 Atmospheric Noise	132
6.2.1 Atmospheric Noise Temperature Ratio	132
6.2.2 Typical Values for Atmospheric Noise	135
6.2.3 Typical Values for Man-Made Noise	135
6.2.4 Values of Noise in the UK	137
6.2.5 Antenna Size and Atmospheric Noise Term, F_k .	138
6.2.5.1 Example: high noise conditions	139
6.2.5.2 Example: low noise conditions	140
6.2.6 Noise Performance: Example	140
6.2.6.1 Typical Antenna	140
6.2.6.2 Received Noise Voltage	140
6.2.6.3 Received Signal Voltage	141
6.2.6.4 Adequate Size of Antenna	142
6.2.7 Up-link v. Down-link Operation	142
6.2.7.1 Range: Up-link Operation	142
6.2.7.2 Range: Down-link Operation	143
6.2.8 Optimum Frequency in the Presence of Noise	144
6.3 Noise Factor of a Receiver Antenna	148
6.3.1 Parallel-tuned Antenna	149
6.3.1.1 Noise Factor at Resonance	149
6.3.1.2 Noise Factor Away From Resonance	149
6.3.1.3 Noise Factor of Damped Antenna	149
6.3.2 Series-tuned Antenna	150
6.3.3 Untuned Antenna	150
6.3.4 Summary and Observations	151
6.3.4.1 Parallel-Tuned Antenna	151
6.3.4.2 Series-Tuned Antenna	151
6.3.4.3 Untuned Antenna	152
6.3.4.4 Example	153
6.3.4.5 Summary	153
6.3.5 Comparison with Atmospheric Noise	154
6.3.6 Increasing Bandwidth by Equalisation	154
6.3.6.1 Post-amplifier filtering	154
6.3.6.2 Magnetic Field Feedback	155
6.3.7 Increasing Bandwidth by Antenna Design	155
6.4 Amplifier Noise	155
6.4.1 Discrete Devices	155
6.4.2 Integrated Circuits	156
6.5 Interference	157
6.5.1 Shielding	158
6.5.2 Using a Large Transmitter	158
6.5.3 Positioning to minimise noise	158
6.5.4 Noise-Cancelling Antennas	159
6.5.4.1 Field Gradient Method	159
6.5.4.2 E-field Cancellation	159
6.5.5 Spatial Filtering	160
6.5.6 Adjacent Channel Cancellation	161
6.5.6.1 Carrier extraction and demodulation	161

6.5.6.2	Interference cancellation	161
6.5.7	Quadrature Phase Space	163
6.5.8	Speech Processing	164
6.6	Concluding Remarks	164
6.7	References	165
7	CHANNEL SOUNDING	167
7.1	Principles of Channel Sounding	167
7.1.1	Purposes	167
7.1.1.1	Frequency Stepping	168
7.1.1.2	Wideband Methods	169
7.1.1.3	Signal Averaging	169
7.2	Outline of Proposed Method	170
7.2.1	Sequence Design	170
7.2.1.1	M+1 Sequence	170
7.2.1.2	Synchronisation	171
7.2.2	Frequency Range	172
7.2.3	Initial specification	173
7.2.3.1	Transmitter	173
7.2.3.2	Receiver	173
7.3	Design Calculations	173
7.3.1	Antenna Design	173
7.3.2	Choice of Sounding Sequence	173
7.4	Integrating Sounding Into a System	175
7.5	Concluding Remarks	177
7.6	References	177
8	SEQUENCE DESIGN FOR SYSTEM IDENTIFICATION	178
8.1	Inverse Sequences	178
8.1.1	Derivation of an Inverse Sequence	179
8.1.2	Verification of Result	179
8.1.3	Observations	181
8.1.3.1	On the sum of coefficients	181
8.1.3.2	On the existence of an inverse	182
8.1.3.3	When the mean of f is zero	182
8.2	Noise Factor	183
8.2.1	Derivation of Signal/Noise Ratio	183
8.2.2	Analogy to Matched Filters	184
8.2.3	Investigation of Noise Factor Expression	185
8.2.4	Sequences with imperfect ACF	186
8.3	System Synchronisation	186
8.3.1	Code-Locked Loop	186
8.3.2	An improved code detector	187
8.3.3	Ripple caused by zeroes in R	189
8.3.3.1	Example: Ripple in control signal	189
8.3.3.2	A zero dead-band early/late control signal	191
8.3.4	Noise factor of a code-locked loop	192
8.4	Orthogonal Sequences	194
8.4.1	Cases of Correlation	194
8.4.1.1	Sequences Completely Uncorrelated	194

8.4.1.2 Correlation is time-shifted delta function	194
8.4.1.3 Generalisation	195
8.4.2 Summary of Orthogonality Results	196
8.5 Concluding Remarks	196
8.6 References	196
9 CHANNEL SOUNDER: DESIGN & RESULTS	198
9.1 Design Evolution	198
9.1.1 Data Capture	198
9.1.2 Data Storage & Transmission	199
9.1.3 Microcontroller	200
9.1.4 Receiver synchronisation	201
9.2 Final Design	201
9.2.1 Status of Design	201
9.2.1.1 Receiver	201
9.2.1.2 Transmitter	203
9.3 Evaluation of Prototype	203
9.3.1 Programming the TDS controller	203
9.3.2 Data Processing	204
9.3.2.1 A to D Converter	204
9.3.2.2 Fourier Transform & Windowing	204
9.3.3 Antenna Design and Calibration	205
9.4 Analysis of Results	205
9.4.1 Self-Noise	205
9.4.2 Signal Averaging	206
9.4.3 Spectrum Survey: 3 orthogonal axes	208
9.4.3.1 Vertical Field	208
9.4.3.2 Analysing the timebase peak	208
9.4.3.3 E-W Field	209
9.4.3.4 N-S Field	209
9.4.4 A Detailed View of the Spectrum	213
9.5 Concluding Remarks	216
10 CONCLUDING REMARKS & RECOMMENDATIONS FOR FURTHER WORK	217
10.1 Propagation	217
10.2 Antenna Design	218
10.3 Transmitter Design	218
10.4 Noise and Receiver Design	219
10.5 Channel Sounding	219
10.6 Concluding Remarks	220
APPENDICES	
A1 AUTHOR'S PUBLICATIONS	221
A1.1 Peer-reviewed & Published Conference Proceedings	221
A1.2 Other Conference Presentations	221
A1.3 Peer-reviewed Papers	222
A1.4 Papers in Preparation	222
A1.5 Book Chapters	222
A1.6 CREG Articles	222
A1.7 Magazine Articles	222

A2 PROPAGATION	223
A2.1 Calculating Fields using Retarded Potential	223
A2.2 Vector Potential due to Small Circular Loop	225
A2.3 Exact Expression for Skin Depth	226
A2.4 The Low-Frequency Window	227
A2.5 Computer Simulation of a Buried VMD	229
A2.5.1 MatLab Simulation	229
A2.5.2 Inspection of the Sommerfeld Integrand	231
A3 DRIVING AN INDUCTIVE LOAD	234
A3.1 Sinusoidal Signals	235
A3.1.1 Power Transfer Efficiency	235
A3.1.1.1 Observation on matching co-efficient	236
A3.1.2 Load Power Dissipation	237
A3.1.3 Amplifier Power Dissipation	237
A3.2 Square-Wave Signals	237
A3.2.1 Complementary and H-Bridge Output	237
A3.2.2 Push-Pull Output	238
A3.3 Partially-tuned Load	239
A4 RECEIVER NOISE-MATCHING	240
A4.1 Noise Effects of Bias Resistor	240
A4.2 Noise Factor of Parallel-tuned Antenna	240
A4.2.1 Notation	240
A4.2.2 3dB Bandwidth	241
A4.2.3 General Expression	241
A4.2.4 Approximations	242
A4.2.4.1 At Resonance	242
A4.2.4.2 Off-Resonance	242
A4.3 Noise Factor of Series-tuned Antenna	243
A4.3.1 3dB Bandwidth	243
A4.3.2 General Expression	243
A5 SEQUENCE DESIGN	245
A5.1 Fourier Transform Pairs	245
A5.2 Convolution Theorem, and Correlation	245
A5.3 Parseval's Theorem	246
A5.4 Power Spectral Density	246
A6 CAVE LOCATIONS FOR TRIALS	248
A6.1 References	249
A7 CHANNEL SOUNDER DESIGN	250
A7.1 Design Status	250
A7.1.1 Hardware	250
A7.1.2 Software	250
A7.2 TDS 2020 Motherboard	250
A7.2.1 External Interface	251
A7.3 DSP Capture Board	253
A7.3.1 Description of Digital Section	256
A7.3.1.1 Performing the signal averaging	256

A7.3.1.2	Reading the data	257
A7.3.1.3	Testing	258
A7.3.1.4	Control Signals	258
A7.3.2	Description of Analog Section	259
A7.3.2.1	Signal Conditioning	259
A7.3.2.2	Analogue to Digital Conversion	264
A7.4	EPLD Programming	265
A7.4.1	PG1 – data latch	265
A7.4.2	PG2 – gates and control latches	265
A7.4.3	PG3 – timing generation	265
A7.5	References	265
A8	DATA PROCESSING	266
A8.1	Window Functions	266
A8.1.1	Generalised Cosine Windows	266
A8.1.1.1	Sidelobe height	267
A8.1.1.2	3dB and 6dB bandwidth, first zero	267
A8.1.1.3	Noise-Equivalent Bandwidth (NEB)	267
A8.1.1.4	Coherent Gain (CG)	268
A8.1.1.5	Scalloping Loss (SL)	268
A8.1.2	Window Coefficients	268
A8.2	MatLab's DFT Function	269
A8.2.1	Distribution of Coefficients	269
A8.2.2	Signal Power	269
A8.2.3	Sample Numbering	270
A8.3	References	270
A9	PROGRAM LISTINGS ON CD-ROM	271
A9.1	MatLab Programs	271
A9.1.1	Propagation	271
A9.1.2	Atmospheric Noise	271
A9.1.3	Sequence Design for System Identification	272
A9.1.4	Spectrum of M+1 Sequence	272
A9.1.5	Channel Sounder Results	272
A9.1.6	Window Functions	272
A9.2	EPLD Programs	272
A9.3	Spreadsheet Programs	272

Glossary of Symbols and Abbreviations

List of Abbreviations

ACF	auto-correlation function
ADC	analogue to digital converter
AM	amplitude modulation
ASS	absolute signal strength
BCRA	British Cave Research Association
BER	bit error rate
c.s.a.	cross-sectional area
CCF	cross-correlation function
CCIR	Comité Consultatif International des Radiocommunications
CLL	code-locked loop
CMRR	common-mode rejection ratio
CREG	BCRA's Cave Radio & Electronics Group
CRO	Cave Rescue Organisation
DFA	distance and field angle
DFT	discrete Fourier transform
DPSK	differential PSK
DSBSC	double sideband (and) suppressed carrier
DSP	digital signal processor/processing
ELF	extra low frequency, 30–300Hz
ETSI	European Telecommunications Standards Institute
FFT	fast Fourier transform
FLL	frequency-locked loop
FM	frequency modulation
GZ	ground zero
HED	horizontal electric dipole
HF	high frequency, 3–30MHz
HMD	horizontal magnetic dipole (vertical loop)
I	in-phase carrier
IC	integrated circuit
IDFT	inverse discrete Fourier transform
IS	intrinsic safety
ISB	independent sidebands
ITU	International Telecommunication Union
LF	low frequency, 30–300kHz

LSB	lower sideband; least significant bit
MCU	microcomputer/controller unit
MSA	magnetic specific aperture
MSB	most significant bit
MS-DOS	Microsoft® disc-based operating system
NBFM	narrowband FM
NEB	noise-equivalent bandwidth
NNEB	normalised NEB
PA	power amplifier
PC	personal computer
PLL	phase-locked loop
p.p.m.	parts per million
PSD	power spectral density
PSK	phase shift keying
Q	quadrature-phase carrier
QPS	quadrature phase space
r.m.s.	root mean square
RA	Radiocommunications Agency
SNR	signal to noise ratio
SSB	single sideband
SWT	single-wire telephone
TVI	television interference
USB	upper sideband
VED	vertical electric dipole
VFG	vertical field gradient (i.e. gradient of vertical field)
VHF	very high frequency, 30-300MHz
VLF	very low frequency, 3-30kHz
VMD	vertical magnetic dipole (horizontal loop)

List of Selected Mathematical Operators

\otimes	Periodic convolution operator
\oplus	Periodic correlation operator
\wedge	Vector product operator
∇	Vector differential operator
$ F $	Modulus of complex value F
\tilde{F}	A tilde, \sim , denotes the reversal of a sequence of samples, F
F^*	An asterisk, $*$, denotes the complex conjugate of F
$\ x\ $	The ‘matching’ function, $\ x\ = x + 1/x$

$\Rightarrow, \Leftrightarrow, etc$	Implication
$f \leftrightarrow F$	Denotes Fourier transformation of f to F (and <i>vice versa</i>)
\bar{F}	A bar, $\bar{\cdot}$, denotes the mean of a sequence of samples, F
ACF, etc.	Functions; names given in the list of abbreviations

List of Symbols

α	Angle of magnetic field line to horizontal Angle subtended by air-gap
γ	Attenuation coefficient, $\gamma = 1/\delta$
Γ	Gain term used in analysis of tuned antenna
δ	Skin depth, (<i>suffix e</i> – external medium) Loss angle 'delta' function
ΔT	Time interval, e.g. sample period, clock period
ϵ	Electric permittivity or dielectric constant
ϵ_0	Electric permittivity of free space, $\approx 8.854 \times 10^{-12} \text{ F m}^{-1}$
ϵ_r	Relative permittivity
ζ	General usage
η	Surface impedance [Ω] Matching coefficient [dimensionless]
$\hat{\theta}$	Unit vector in spherical polar co-ordinates (r, θ, ϕ)
θ	Angle between z -axis and r vector in (r, θ, ϕ) co-ordinates Bandwidth widening factor of a damped tuned circuit General usage
λ	Wavelength Dummy variable of integration
μ	Magnetic permeability (<i>suffix e</i> – external medium)
μ_0	Magnetic permeability of free space, $= 4\pi \times 10^{-7} \text{ H m}^{-1}$
μ_r	Relative permeability
μ'	Effective permeability
μ'_r	Effective relative permeability
ρ	Mass density Distance from origin in $\theta = \pi/2$ plane for spherical polar co-ordinates (r, θ, ϕ) . Offset of field point from ground zero.
σ	Electrical conductivity (<i>suffix e</i> – external medium)
τ	Volume of a current distribution
$\hat{\phi}$	Unit vector in spherical polar co-ordinates (r, θ, ϕ)
ϕ	Angle between y -axis and projection of r onto $\theta = \pi/2$ in (r, θ, ϕ) co-ordinates Magnetic flux
Φ	Specific Aperture [$\text{m}^2 \Omega^{-1/2}$]

Φ_m	Specific Aperture of a magnetic-cored antenna [$\text{m}^2 \Omega^{-1/2}$]
Φ_e	Specific Length of an electric dipole antenna [$\text{m } \Omega^{-1/2}$]
with prefix:	t – transmitter, r – receiver
Φ_s	Sensitivity of loop antenna [$\text{T Hz}^{-1/2}$]
Ψ	Demagnetising factor
ω	Angular frequency, $\omega \equiv 2\pi f$. (suffix e – external medium)
ω_c	Cut-off frequency (of a filter), characteristic frequency
ω_0	Centre frequency of tuned circuit
Ω	Frequency interval between samples $\equiv \omega / \omega_0$, i.e. normalised frequency
$\overline{\Omega}$	$\equiv \Omega - 1/\Omega$
a	Radius (e.g. of loop, solenoid or radome)
\mathbf{A}	Magnetic vector potential
A	Area, e.g. cross-sectional area of loop
A_c	Capture cross-section, defined here as A_e / G
A_e	Effective aperture (not to be confused with specific aperture Φ)
A_n	Amplifier noise term [dimensionless]
B	Bandwidth
B_s	Signal bandwidth
B, \mathbf{B}	Magnetic Flux Density (scalar, vector), $\mathbf{B} = \mu \mathbf{H}$
B_{sat}	Saturation flux density
c	Speed of light in a vacuum
d	Distance
D	Normalised offset from GZ, $D = \rho/h$ Distance
D_{max}	Maximum distance (i.e. range) of a transmission
D, \mathbf{D}	Electric Flux Density (scalar, vector), $\mathbf{D} = \epsilon \mathbf{E}$
E, \mathbf{E}	Electric Field (scalar, vector) Energy per bit
\hat{E}_n	Electric noise field spectral density [$\text{V m}^{-1} \text{Hz}^{-1/2}$]
f	Frequency
f_0	Clock frequency
f, F	Sequence of samples in time domain, and its DFT
F_a	Atmospheric noise temperature ratio
F_k	Atmospheric noise term [$\Omega^{1/2} \text{m}^{-2}$]
F_n	Noise factor of receiver amplifier [dimensionless]
g	Power amplifier output voltage swing, as fraction of maximum
G	Antenna gain (r.f. terminology)
g, G	Sequence of samples in time domain, and its DFT
h	Depth of source or field point below surface

h_e	Effective height
h, H	Sequence of samples in time domain, and its DFT
H, \mathbf{H}	Magnetic Field (scalar, vector)
\hat{H}_n	Magnetic noise field spectral density [$\text{A m}^{-1} \text{Hz}^{-1/2}$]
\hat{i}_n	Amplifier noise current spectral density [$\text{A Hz}^{-1/2}$]
I, \mathbf{I}	Current (scalar, vector)
$IDFT$	Function: inverse discrete Fourier transform
j	$\sqrt{(-1)}$
J, \mathbf{J}	Current Flux Density (scalar, vector), $\mathbf{J} = \sigma \mathbf{E}$
J_0, J_1	Zero and First order Bessel functions
k	A general constant Boltzmann constant, $\approx 1.381 \times 10^{-23} \text{ J K}^{-1}$
k_0	Wave number $k = 2\pi/\lambda = \omega/c$
k_r	Wave number for free space
	Real part of complex wave number
K	Magnetic current [V]
K_0, K_1	Zero and First order modified Bessel functions
K_Q	Constant used for brevity in derivation of Q -factor
ℓ	Length, e.g. of electric dipole antenna or solenoid
with suffix:	c – of ferrite core of a solenoid, w – of winding on a solenoid
\mathbf{l}	Vector length element
m	Shape factor of solenoid, $m = \ell / 2a$ 'matching' term, a function of η and Q Integer, counting samples in a sequence
m_d	Magnetic Dipole Moment [Am^2]
M	Length of a maximal-length sequence
M_d	Demagnetising factor (Burrows' definition)
M	Mass (e.g. of wire on an antenna)
with suffix:	c – of ferrite core of a solenoid, w – of winding on a solenoid
n	Integer, counting number of samples in a sequence noise voltage or power
n_0	Gaussian noise power spectral density [W Hz^{-1}]
N	Number of turns Noise voltage Integer
p_d	Electric Dipole Moment [A m]
P	Horizontal field coefficient [dimensionless]
with suffix:	Power (specifically, power dissipation in an antenna) d – Power density of electromagnetic field e – Power drawn from source of energy (e.g. battery drain) l – Power dissipation in a loop antenna p – Power dissipation in a power amplifier
\hat{P}_n	Amplifier noise power spectral density [W Hz^{-1}]

Q	Vertical field coefficient [dimensionless]
Q -factor	
Q_0	Q -factor, measured at ω_0 or f_0
Q_c	Q -factor of a capacitor
Q_L	Q -factor of an inductor
q	Electric charge
r, R	Sequence of samples in time domain, and its DFT
\hat{r}	Unit vector in spherical polar co-ordinates (r, θ, ϕ)
r, \mathbf{r}	Radius (e.g. of loop, solenoid)
\mathbf{r}'	Distance (as scalar or vector) from origin in (r, θ, ϕ) co-ordinates
r'	Distance (vector) from origin to current element
R	Distance from current element to field point
	Resistance of circuit elements
	Resistance of induction loop antenna
with suffix:	b – bias, d – damping, e – external, l – load, n – noise, r – radiation, p – PA output driver source
S	Signal voltage or power
t	Thickness of winding on a solenoid
	Time
T	Time interval between samples
	Normalised depth of transmitter, $T = h/\delta$
\mathcal{T}	Definition of normalised depth used by J.R. Wait, $\mathcal{T} = \sqrt{2} \cdot T$
T_c	Value of normalised depth T associated with cut-off frequency ω_c
T	Temperature [K]
T_0	Ambient temperature [K], 290K
T_a	Atmospheric noise temperature [K]
U	Voltage
with suffix:	X – reactive, l – of load element, p – of power amplifier
v	Velocity of wave propagation
	Ratio of winding cross-section to antenna cross-section for loop
\hat{v}_n	Amplifier noise voltage spectral density [V Hz $^{-1/2}$]
V_c	Core volume, for ferrite-cored solenoid
x	General quantity, general distance
X	Reactance of circuit elements
$\hat{\mathbf{z}}$	Unit vector in cylindrical polar co-ordinates (r, θ, z)
z	Elevation of source or field point above surface
Z	Normalised elevation of field point, $Z = z/h$
	Impedance of circuit elements
	Wave impedance
Z_0	Wave impedance of free space, $Z_0 = \sqrt{(\mu_0/\epsilon_0)}$

List of Figures

Figure	Title	Page
Figure 1-1	Field from a small magnetic dipole.	9
Figure 1-2	A comparison of up-link and down-link SNR.	18
Figure 2-1	Magnetic dipole field v. distance for a vacuum.	26
Figure 2-2	Magnetic dipole, showing different direction of near and far-field lines.	27
Figure 2-3	Magnetic dipole field v. distance for a vacuum (expanded view).	28
Figure 2-4	Skin depth, as a function of conductivity and frequency.	33
Figure 2-5	Magnetic dipole field v. distance for a conducting medium.	34
Figure 2-6	Magnetic dipole field v. distance for a vacuum (expanded view).	35
Figure 2-7	Ratio of coaxial to coplanar field v. distance.	35
Figure 2-8	Some approximations to the full-conducting-space model.	37
Figure 2-9	Optimum Frequency Window.	41
Figure 2-10	Reflections of elemental dipoles in a conducting half-space.	44
Figure 2-11	Eight basic alignments of transmitting and receiving induction loops.	45
Figure 2-12	Classification of communications systems, based on signal path.	46
Figure 2-13	Comparison of full-plane and half-plane models.	50
Figure 2-14	Q -field v. normalised offset D .	51
Figure 2-15	Q -field v. normalised offset D (detail).	51
Figure 2-16	Q -field v. normalised offset D (region near $T = 6$).	51
Figure 2-17	P -field v. normalised offset D .	52
Figure 2-18	Induced voltage v. frequency, co-axial antennas.	54
Figure 2-19	Induced voltage v. frequency, comparison of models.	54
Figure 3-1	Specific Aperture of different shaped antenna loops.	66
Figure 4-1	A simple model of a J -field ‘earth-current’ antenna.	83
Figure 4-2	Ferrite rod antenna with winding on central section.	87
Figure 4-3	Comparison of antenna arrays.	90
Figure 4-4	Fringing effect, or ‘demagnetisation’ of a solenoid.	93
Figure 4-5	Alignments of multiple ferrite rods.	98
Figure 4-6	B - H hysteresis curve for a permanent magnet.	107
Figure 5-1	Class-B/C push-pull power amplifier driving a resistive load	114
Figure 5-2	Class-B/C amplifier driving an inductive load with current feedback.	118
Figure 5-3	Half H-bridge complementary output for digital amplifier.	120
Figure 6-1	Expected value of world-wide atmospheric noise below 10kHz.	136
Figure 6-2	Expected noise temperature from 10kHz to 10MHz.	136

Figure 6-3	Likely range of noise terms Fa and Fk.	139
Figure 6-4	A comparison of up-link and down-link SNR.	143
Figure 6-5	Variation of SNR with frequency and atmospheric noise temperature profile.	145
Figure 6-6	Variation of SNR with frequency for different depths.	147
Figure 6-7	Tuned and untuned antenna configurations.	148
Figure 6-8	Receiver architectures for interference rejection.	162
Figure 7-1	Block diagram showing intended operation of the channel sounder.	170
Figure 7-2	Spectrum and antenna current due to 256-bit sounding sequence.	174
Figure 8-1	System configuration for identification using test sequence.	179
Figure 8-2	Auto-correlation of a bipolar version of the sequence listed in Table 8-1.	180
Figure 8-3	Plot of the inverse sequence listed in Table 8-2.	181
Figure 8-4	A traditional code-locked loop.	187
Figure 8-5	A code-locked loop control signal for with no dead-band.	188
Figure 8-6	Code-locked loop with single multi-valued cross-correlation.	189
Figure 8-7	Control signal for example in §8.3.3.1.	190
Figure 8-8	Inverse sequence needed to achieve the conventional CLL ‘S-curve’ control signal.	191
Figure 8-9	Inverse sequence needed to achieve a zero dead-band control signal.	192
Figure 9-1	NW-SE elevation of White Scar Cave at grid reference SD 713746.	199
Figure 9-2	Block diagram of prototype channel sounder.	202
Figure 9-3	Self-noise recorded with shorted input.	207
Figure 9-4	Self-noise with signal averaging of 100 datasets.	207
Figure 9-5	Enlargement of Frequency spectrum.	207
Figure 9-6	Vertical field. Time and frequency spectrum.	211
Figure 9-7	Frequency spectrum, vertical field, showing TV timebase.	211
Figure 9-8	Frequency spectrum, vertical field, showing mains harmonics.	211
Figure 9-9	Horizontal E-W field. Time and frequency spectrum.	212
Figure 9-10	Horizontal N-S field. Time and frequency spectrum.	212
Figure 9-11	Horizontal N-S field. Frequency spectrum 40-100kHz.	212
Figure 9-12	Horizontal N-S field. Frequency spectrum 0-50kHz.	213
Figure 9-13	Horizontal N-S field. Frequency spectrum 50-100kHz.	213
Figure 9-14	Horizontal N-S field. Frequency spectrum 100-150kHz.	214
Figure 9-15	Horizontal N-S field. Frequency spectrum 150-200kHz.	214
Figure 9-16	Horizontal N-S field. Frequency spectrum 200-250kHz.	215
Figure A2-1	Retarded potential.	224
Figure A2-2	Retarded potential due to a current loop.	225

Figure A2-3	Low-frequency window.	228
Figure A2-4	Integrand of Sommerfeld integral.	231
Figure A2-5	as Figure A2-4 with parameters as given.	232
Figure A2-6	as Figure A2-4 with parameters as given.	232
Figure A2-7	as Figure A2-4 with parameters as given.	232
Figure A3-1	Class-B/C amplifier driving an inductive load	234
Figure A7-1	TDS Microcontroller and 500Mbyte hard disc card.	252
Figure A7-2	Interconnection of channel sounder modules.	252
Figure A7-3	Circuit diagram of microcontroller motherboard.	254
Figure A7-4	Assembly diagram of microcontroller motherboard.	255
Figure A7-5	Channel sounder DSP board; block diagram.	256
Figure A7-6	Circuit diagram of DSP/capture board, sheet 1 of 3.	260
Figure A7-7	Circuit diagram of DSP/capture board, sheet 2 of 3.	261
Figure A7-8	Circuit diagram of DSP/capture board, sheet 3 of 3.	262
Figure A7-9	Assembly diagram (un-annotated) of DSP/capture board..	263
Figure A7-10	Instrumentation amplifier.	264
Figure A8-1	DFT spectral response and the effect of a window function.	266

List of Tables

<i>Table</i>	<i>Title</i>	<i>Page</i>
Table 1-1	Demonstration of optimum frequency effect.	19
Table 2-1	Definition of skin depth in good and poor conductors.	31
Table 2-2	Skin depth, as a function of conductivity and frequency.	32
Table 2-3	Field classification based on distance of field point.	38
Table 2-4	Field classification based on conductivity of ground.	39
Table 2-5	Bandwidth of optimum frequency window.	41
Table 3-1	Typical parameters for air-cored induction loops.	65
Table 3-2	The merits of different antenna materials.	67
Table 4-1	<i>H</i> , <i>E</i> and <i>J</i> fields - relationship between field strength and flux density.	77
Table 4-2	Comparison of electric and magnetic dipole systems.	80
Table 4-3	Change in specific aperture for five configurations of <i>N</i> rods.	98
Table 4-4	Comparison of typical air-core loops.	100
Table 4-5	Comparison of different alignments of ferrite-cored solenoid antennas.	100

Table 4-6	Comparison of magnetic materials.	109
Table 5-1	Comparison of tuned v. untuned antennas.	126
Table 6-1	Likely range of Atmospheric Noise Temperature Ratio, F_a .	137
Table 6-2	Man-made noise level in different environments.	137
Table 6-3	Seasonal atmospheric noise temperature in southern England.	138
Table 6-4	Likely range of noise term F_k .	139
Table 6-5	Example: receiver antenna voltage due to signal and sources of noise.	142
Table 6-6	Amplifier noise-matching term for different configurations of antenna.	152
Table 6-7	Noise characteristics of some operational amplifiers.	156
Table 7-1	Bandwidth v. Q -factor for first-order tuned circuit.	172
Table 8-1	256-bit test sequence $f(nT)$.	180
Table 8-2	Inverse sequence, rounded to 2 decimal places.	181
Table 8-3	Generator sequences for control signal in example in §8.3.3.1.	190
Table 8-4	Data for Figure 8-8.	191
Table 8-5	Data for Figure 8-9.	192
Table 9-1	Values of mains harmonics, from Figure 9-8.	209
Table A2-1	Bandwidth of low frequency window.	229
Table A2-2	MatLab program listing for function 'sommerfeld'.	230
Table A2-3	MatLab program listing for function 'PQintegral'.	230
Table A8-1	Parameters for a range of generalised cosine windows.	267

1 Sub-Surface Communications

This thesis is a study of sub-surface radio communication. In this, we distinguish between 'wireless' systems and a variety of systems that utilise a cable. The distinction is not always easy to make – the example of a leaky-feeder antenna being a system that depends both on 'cable' and 'wireless' transmission. We can, of course, distinguish between systems that operate *within* tunnels (such as the leaky-feeder), and those that depend on through-rock penetration but, even then, there are some hybrid systems that are difficult to classify. The focus of this thesis is on subterranean communication that depends primarily on the penetration of electromagnetic fields through the ground.

1.1 Uses for Sub-Surface Communications

In addition to uses in the mining industry, sub-surface communication is used for bore-hole telemetry in the geophysical and oil industries, for pipeline location by water and gas services and is increasingly used by cavers and pot-holers. Uses for caving include rescue communications, radiolocation for surveying purposes and telemetry. We shall limit these studies to mine and cave applications. But despite the apparent similarity, the system requirements for cave and mine communications are somewhat different.

1.1.1 Mine Applications

For mine applications the equipment has to meet stringent *intrinsic safety* (IS) requirements, which limit the power that can be used, the choice of electronic components and even the material from which the equipment housing is constructed. As an example, a small, tuned induction loop transmitter is not feasible because the resonant voltage would exceed IS constraints. Neither can a radio be battery operated if this would require that the batteries be changed underground, unless special precautions are taken. For this reason, one induction radio device designed for mines usage features a hand-cranked generator.

Early research into mine communications concluded that it was not practical due to the large power required but a spate of mining accidents lead to an intensive programme of research by the US Bureau of Mines in the 1960s. Induction radio systems are now in use in many mines; see for example, review articles (Austin, 1973; 1987), (Murphy, 1978) and (Durkin, 1984). Murphy describes a small pager system to alert miners to danger. A more sophisticated location

system using a binary maximal length sequence and a coherent receiver is described in (Ristenbatt, Holland-Moritz and Metzger, 1988). The problem of locating trapped miners using body-worn beacon transmitters is discussed in (Nessler, 1981).

1.1.2 Cave and Geophysical Applications

Cave communication systems differ from those used in mines because there is usually a requirement for a small, portable antenna and a low power consumption, whereas a mining application can use a larger, more powerful, fixed antenna. In addition to speech communication, the equipment is used for radio-location and ground mapping – both in caving and civil engineering applications. Radiolocation beacons make use of the predictable behaviour of the near-field lines, and so accurate results are usually only attainable over short distances. Geophysical applications include the remote measurement of ground conductivity by analysing the phase of the induced fields from an induction loop.

Additionally, for cave applications, equipment has to be rugged and waterproof – probably more so than for use in mines. A particular problem is that caves often contain very fine silt that can coat moving parts such as switches and connectors and then set solid. The speleological environment never requires equipment to be intrinsically safe and so there is greater scope for experimentation with cave communications systems.

1.2 Features of Sub-surface Radio

The main difficulty we face in using sub-surface ‘wireless’ communications is that electromagnetic fields are attenuated as they penetrate a conducting medium. Low frequencies are attenuated less than high frequencies and so it is usual, for speech communication, to use a frequency in the LF band (30–300kHz). This, however, leads to inefficiency in the small, portable antennas we would like to use. Within this frequency range, we will show that there is an optimum frequency that depends on the transmission distance as well as the conductivity of the ground and, for short-range communications in dry rock, much higher frequencies may be useable.

Short-range subterranean radio operates largely in the near-field zone where induction is a more significant effect than radiation. Conversely, the ELF (30–300Hz) systems that have been developed for submarine communications rely on far-field operation, with the propagation involving a long-distance path in air – at 300Hz the transition zone occurs at a distance of about 160km in air, perhaps 300m in rock, but only 15m in seawater. Submarine communication therefore involves quite different principles and will not be considered in this thesis.

1.2.1 Mine Communications

(Austin, 1973; 1987) describes an SSB radio operating at 335kHz in South African mines. In the UK, induction radio equipment operate by the Mines Rescue Service operates at 30-50kHz; equipment for use by cavers has tended to use an 87.5kHz carrier, although some equipment has been built to operate at 100–125kHz. In the UK the Radiocommunications Agency, which regulates the use of the radio spectrum, limits inductive communications to 185kHz, (RA, 1988) with the limitation on field strength being specified by the European Telecommunications Standards Institute (ETSI, 1994). Baseband audio equipment is no longer used, but beacons to aid cave surveying, and which operate in the region of 1-3kHz, are common.

1.2.2 Cave Communications

Although the main commercial use of sub-surface radio is in the mining industry, we are particularly interested in its use by cavers and pot-holers for rescue purposes. With that in mind, it is appropriate to review the current ‘state of the art’ of cave communications.

On deep or wet pitches, communicating by means of whistle-blasts is an established practice and, in many situations, this or perhaps a loud voice will be sufficient. However, for cave rescue, expedition management or the execution of projects such as photography in a large chamber, communication by radio or telephone is often essential.

Possibly the first use of a telephone in a cave was during the exploration of Lamb Leer Cavern (in the UK’s Mendip hills) in 1880 (Williams, 1995). Despite the disadvantage of having to lay a cable, telephones continue to be used because they are simple and robust. A variety of other cable-based systems exist – single-wire telephone, ‘guide wire’ radio, and optical fibre. True ‘wireless’ communication is difficult because rock, being conductive, absorbs radio waves. For line-of-sight work within cave passages, high-frequency (HF) walkie-talkies and CB radios are used but magnetic induction equipment is usually necessary for communication through the rock itself. The technique of ‘earth current injection’ was developed during the First World War (1914-18) but is rarely used now, although an enhancement, using a low frequency (LF) carrier, is the basis for the latest high-performance induction radio / earth-current ‘hybrids’ now used in the UK and Europe.

The single-wire telephone (SWT), also known as an earth-return telephone, uses the conductivity of the ground to provide a return path for the current. The obvious advantage is that only half the weight of cable has to be carried. The devices used by cavers feature electronic amplification and will often operate without any specific earthing other than through the operator’s body. SWTs are cheap, rugged, easy to build by amateurs, and are preferred to

the traditional army field telephone. They are popular with cave expeditions and rescue groups throughout the world, and are also convenient for providing flood-warnings to cavers.

If, instead of audio, we pass a LF radio signal along a SWT cable it will work without any earthing, because there is sufficient capacitive coupling from the operator's body to earth. Additionally, the handset can be coupled to the cable capacitively or inductively resulting in more versatile equipment. Commercial equipment based on this principle is used for mines rescue. HF radio can also be guided along a cable placed in close proximity although the principle of operation is quite different. For commercial use at VHF, 'leaky-feeder' cable presents yet another mode of operation. Confusingly, all these systems are referred to as 'guide wire radio'.

Rock is electrically conductive, which causes it to absorb radio waves, so HF radio is mostly limited to use within cave passages. Rescue teams use walkie-talkies (27 or 49MHz) to communicate on pitches while manhandling an accident victim in a stretcher. Photographer Gavin Newman has used a modified CB radio to co-ordinate photography in Illu River Cave, Irian Jaya. He also found the radios useful to warn of impending flooding in this huge cave system (Newman, 1993). HF radio will penetrate a few metres into rock and can be used for locating dig points on the surface although this normally requires specialised equipment. In 2001 a CB radio was used to verify the surface location of the top of the Titan shaft in Peak Cavern, Derbyshire. Radio waves can travel for some distance along mineralised fault lines or dry well-jointed rock so, occasionally, quite deep penetration of a cave is possible.

Low frequencies are attenuated less than high frequencies – long wave broadcasts such as BBC Radio 4 (198kHz) can be detected at the bottom of deep caves. Unfortunately, detecting a radio signal is easier than transmitting it – a 200kHz signal would require an electric half-wave antenna some 750m long, which is normally only feasible for mining installations. From the early 1960s cavers began to experiment with loop antennas. Because of their small size (typically 1m) these radiate very little 'true' radio energy, and mainly generate a magnetic field which couples to the receiver loop by the principle of magnetic induction, and operates over a limited range of just a few hundred metres. Experimental induction systems have covered 27–185 kHz, (i.e. most of the long-wave broadcast band), and most have used single-sideband (SSB) operation. The most notable designs have been R. Mackin's commercially-produced Molefone, dating from 1979, which has seen widespread use in the UK, and I. Drummond's CB Transverter, used mainly in the US and Canada (Gibson, 1999).

Two of the latest designs are J. Hey's HeyPhone and the French 'Nicola System' designed by G. Naylor. Both operate at 87.5kHz SSB and are published designs (Anon., 2002; Naylor, 1999). They are intended to work with induction loops or earth-current antennas, thus reviving

the early-1900s trench-communications technique. With earth electrodes the Nicola system has been demonstrated to work through 1000m of rock. In 2001, working with the Cave Radio & Electronics Group, the British Cave Rescue Council issued over fifty HeyPhones to the UK's cave rescue teams (Bedford, 2001). This is the first time that a national organisation has systematically equipped its rescue teams with such advanced equipment. The South Wales team soon had occasion to test their HeyPhones. A caver dislocated his shoulder some 5km into Daren Cilau – a formidable cave protected by a daunting 520m entrance crawl and requiring an underground camp to explore the far reaches. Whilst the injured caver slowly made his way to the camp, two of his companions made a strenuous exit and raised the alarm. Some 28 hours later, the medical team was able to communicate with the surface to report that the casualty had been attended to, and would not require a stretcher carry (which, it had been estimated, would take up to three days to complete). Use of the HeyPhones (and of a SWT in the entrance crawl) considerably eased the logistics of this rescue, which involved 63 personnel (Anon., 2001).

Cave radios have been connected, experimentally, to mobile telephones and HF radios on the surface, and messages relayed automatically over long distances (Bedford and Gibson, 1995; Naylor, 2000). System development seems likely to continue, with the aim of providing extra functionality, for example medical data logging, text messaging, and digital image transmission.

1.3 Aims of this Study

The features that characterise through-rock systems are

- i) Communications through conductive media over distances of only a few skin depths.
- ii) Small, portable antennas (large fixed installations have been used in the mining industry).
- iii) Restricted choice of frequency (interference, noise, other primary channel users).
- iv) Optimum frequency and orientation of antenna depends on many external factors, including transmission distance.

Within the above framework we aim to investigate the component parts of a sub-surface communications system, and present guidelines that will aid the design of an 'optimum' system. Such a system could be described as having a high performance for a low cost; and our interpretation of these parameters may depend on circumstances. Two clear indicators of 'performance' are the transmission range (at a cost of antenna size and portability) and transmission time (at a cost of power consumption). Other performance factors such as reliability and security of communications are beyond the scope of this thesis, which will

concentrate on the antenna design and on a study of sub-surface propagation of electromagnetic fields.

Propagation in a conductive medium, although well studied, can still be difficult to describe succinctly. Channel sounding, as well as providing the practical means to determine the optimum operation of a communication system, will hopefully allow us to test the theory and to develop some simple and instructive models of the propagation.

1.4 Sub-Surface Penetration of Electromagnetic Fields

1.4.1 Propagation

Sub-surface electromagnetic communication takes place in a dissipative medium. The ground has a finite electrical conductivity, which can vary from typically 0.1mS/m for dry rock, to 10mS/m for wet rock; and in the case of submarine communications in seawater it is typically 4S/m (Burrows, 1978). Electromagnetic waves penetrating such a medium obey the diffusion equation $\nabla^2 \mathbf{H} = \mu\sigma \partial \mathbf{H} / \partial t$ rather than the wave equation $\nabla^2 \mathbf{H} = \mu\epsilon \partial^2 \mathbf{H} / \partial t^2$, with the field decaying exponentially with distance into the medium and the degree of decay of the field expressed by the *skin depth*, δ . In a uniform conducting medium a plane wave is attenuated exponentially with distance x , at a rate of $\exp(-x/\delta)$, or -8.7dB per skin depth. In a good conductor, the skin depth is inversely proportional to the square-root of the frequency and so, although VHF signals have been used for short-range communications through dry rock, it is usually necessary to use frequencies in the LF (30–300kHz) or VLF (3–30kHz) bands. Voice-band frequencies have also been used – for beacons as well as baseband audio. Submarine communications may exploit require even lower frequencies, in the ELF band (30–300Hz) and below.

Both the electric and magnetic fields are attenuated in accordance with the ‘skin depth’ behaviour outlined above but there is a further phenomenon to consider. An electromagnetic wave, incident on the boundary between air and a conducting medium will undergo reflection and refraction. Because the medium is highly conductive¹, most of the electric field is reflected and the transmitted portion undergoes an attenuation of $20 \log_{10} (\sigma / \epsilon_0 \omega)$, where ϵ_0 is the electric permittivity of free space. Typically, at 100kHz, with $\sigma = 1\text{mS/m}$, the electric field attenuation across the boundary is 45dB and it will be greater at lower frequencies. For this reason, communications systems that rely on receiving the electric field are not usually feasible,

¹ with $\sigma / \omega\epsilon \gg 1$, where σ is the conductivity, ω is the angular frequency and ϵ is the electric permittivity of the medium.

although a more comprehensive analysis is required, which allows for the performance of the transmitter antenna. Such a discussion takes place in chapter 4.

1.4.2 Antennas

An obvious implication of using a low frequency is that a resonant, radiating antenna needs to be large – at 100kHz a half-wave dipole is 1500m in length. For fixed installations used in submarine or commercial mining applications this is not a problem – at least, not for the surface-to-subsurface link. However, where portable antennas are required, the radiation efficiency will be low if the antenna is much smaller than its resonant size. Fortunately, this problem is offset by the fact that communication may be required over just a short range. If the range is within the *near-field* of the transmitter then an efficient radiator is not required because the main mode of operation will be by inductive coupling (for a magnetic dipole antenna) or electrostatic coupling (for an electric dipole antenna).

Two fundamental forms of antenna are the electric and the magnetic dipole. When self-resonant, the structures appear as resistive loads. However, when the antenna is much smaller than the self-resonant size, the electric dipole presents a low capacitance load, whilst the magnetic dipole presents a high inductance load. In this situation the antenna requires tuning in order to drive it in the most efficient manner and this can result in a dangerously high terminal voltage. But the capacitance of 1m electric dipole, when it is much smaller than a wavelength in size, is only a few pF and so stray coupling will affect its performance, both as a transmitter and receiver. This presents us with a second reason for preferring to utilise magnetic fields rather than electric fields for sub-surface communications.

In chapter 4 we will discuss some other possible antennas but the magnetic dipole or induction loop antenna remains the preferred choice for portable communications.

1.4.3 Near and Far Fields; Transition Distance

We noted, above, that the short range of communications meant that we would probably be operating in the near field. The distinction between near and far-fields will be discussed in chapter 2 but, in summary, the near field represents a zone where the field has the characteristics of a static dipole, and decays with the inverse cube law. Whereas the far-field represents the zone where radiation is predominant and the field – in air – decays with an inverse linear law.

However, far-field effects cannot be disregarded entirely because, in a conducting medium, the *transition zone* from near to far-field contracts from $\lambda/2\pi$, in air, to the skin depth, δ . It is likely that cave communications will be operating within this zone. For mining applications,

transition-zone operation may be inevitable, because many minerals are conductive, and mine workings are often extensive. Submarine communications are usually far-field in nature, and the most direct path for the signal can involve an 'up and over' route through the air because seawater is so highly conductive that skin-depth attenuation is severe. Each decade of the radio spectrum demonstrates its own particular characteristics of propagation and usage, and this is as true of ELF and LF waves as it is of the higher frequencies. In this thesis we shall concentrate on voice-band and LF communications; the particular characteristics of submarine ELF signals will not be discussed further.

In a conducting medium the radiation field is minuscule because the energy is absorbed by the conducting medium and, strictly, it would be incorrect to describe sub-surface communications as taking place "in the far-field". Where we use this terminology, it should be taken to mean the zone where simple near-field characteristics are modified by secondary induced fields.

When power is supplied to an electrically small antenna, only a very small proportion (the *radiation field* or *far-field*) is radiated. Most of the power is simply dissipated as heat in the antenna but, at the same time, it generates large fields in the vicinity of the antenna. For an isolated antenna in a vacuum, these *near-fields* do not convey energy because they are in phase quadrature, and so the product $\mathbf{E} \wedge \mathbf{H}$, integrated over a sphere of a fixed radius, tends to zero over time. The existence of the transition zone between the near and far fields will be demonstrated in chapter 2. However we can begin to appreciate its existence if we consider that the velocity of the field is limited to the speed of light. If we accept that the field has to collapse to zero between half-cycles of current then a heuristic argument suggests that this will not be possible for a field point that is distant from the source. We can imagine, perhaps, that the field lines must therefore become detached from the source and propagate as radiation. To a first approximation the critical detachment distance would be $\lambda/4$, since at this distance it would take a time equal to a half cycle of oscillation for the field lines to collapse and re-form in anti-phase. With some considerable artistic licence, it is possible to represent this on a diagram – *Figure 1-1* below.

Clearly there must be a gradual trend in the form of the field throughout the *transition zone* between near- and far-fields, rather than a sudden change at the *transition distance*. Depending on the purpose of the analysis, it is possible to define this distance in several ways; one convenient definition, as we shall show later, places it at $\lambda/2\pi$. If the field point is sufficiently closer than this, the predominant fields follow the form of electrostatic and induction fields – i.e. they are *quasi-static* in nature, and convey no radiated power. Conversely, if the field point is sufficiently far away then the radiated fields will predominate. For a range of distances either

side of the transition distance, neither the near-field nor the far-field are an adequate approximation to the field behaviour. Within the transition zone, the interaction of the near and far fields makes the fields complicated to describe – especially on a static diagram – because there is an element of elliptical polarisation to the field (that is, the direction of the field line varies depending on which part of the cycle is being depicted). This is not just a theoretical problem, sub-surface communication and radiolocation are both affected by the field characteristics in the transition zone, although the conducting medium itself has a further effect on the field distribution.

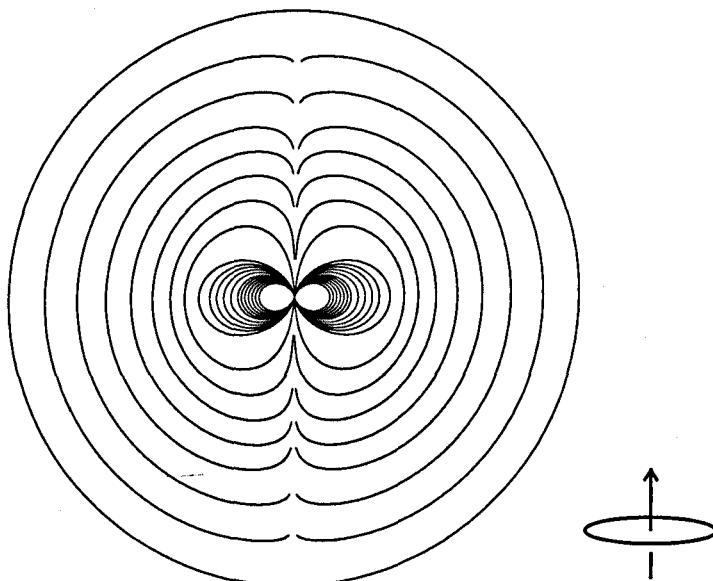


Figure 1-1 : Field from a small magnetic dipole.

The plane of the dipole is perpendicular to the page, with its axis vertical (see inset, right). The inner region shows the induction or near-field, the outer region represents the radiation or far-field. A transition zone separates the regions. Along the axis there is no far-field, so it is not possible to assign a field direction to this region, (hence the apparent break in the field lines). The scale on this figure is considerably non-linear.

1.4.3.1 *The Effect of a Conducting Medium*

Although the near field does not convey energy in a vacuum, if it is intercepted by a conductor, or penetrates a conducting medium, then energy dissipation will take place. The resulting current flow (the ‘eddy current’ in the case of a conducting medium) will generate a **secondary field** which will be out of phase, and possibly in a different direction to the primary field. This field, in turn, affects the source, changing the apparent resistance of the antenna. The effect is exactly that of a loosely-coupled transformer, and the apparent load resistance is known as the **external resistance**. In certain regions of a conducting medium, the secondary field will result in the receiver detecting an elliptically polarised signal. This affects the use of induction fields for positioning and radiolocation (Gibson, 1996a), though it has an advantage

too, in that geophysicists can use this signal as a way of remotely measuring ground conductivity. (Keller and Frischknecht, 1966) §6:38, (Wait, 1982) §3. A report of practical measurements of conductivity in US mines is given in (Durkin, 1991) and for a practical account of this feature, used by cavers, see (Pease, 1991).

1.4.4 Modelling the Field Equations

In the early days of sub-surface induction radio experiments, almost 40 years ago, it was assumed by amateur cave experimenters that sub-surface radio operated purely on the principle of induction, and that the field was quasi-static in nature. It is, perhaps, surprising that this assumption should have made because even an elementary analysis shows this not to be the case. However, for base-band audio, a quasi-static approximation would often have been adequate, and for radiolocation (usually at < 4kHz) this approximation is reasonably good for distances less than around a third of a skin depth.

1.4.4.1 *Fields in a Uniform Infinite Conducting Medium*

To describe the fields in a uniform, infinite conductor we substitute the parameters of the medium into the standard field equations. This is trivial to do, and results in a good basic model for communications over distances that are small compared with the extent of the medium. However, if the medium is not infinite in extent we will need to impose boundary conditions for its evaluation, and this results in a more complex model.

1.4.4.2 *Fields in a Uniform Conducting Half-Plane*

We could surmise that a more complex model than the above would be required for the dual conditions that

- i) Communication is over a distance that is significant in comparison to a skin depth, and
- ii) Either a) the communications path crosses the air-rock boundary
or b) the transmitter or receiver is within a skin-depth of the air-rock interface.

In the latter case, if the transmitter and receiver were both in air, we would expect induced ground currents to influence the propagation. Likewise, if the transmitter and receiver were both buried, we could expect a significant proportion of the field to travel by an 'up over and down' path which would suffer less attenuation than a direct path through the rock. Boundary conditions also need to be applied to the situation of an antenna in a buried radome and if we wish to consider the effect of stratified ground consisting of layers of rock of different conductivity.

The general solution to this problem is complicated, although it often possible to simplify the treatment. For example, we have already discussed the trivial case where the field point is sufficiently close to the source that we can assume a quasi-static approximation. However, as

we have seen, sub-surface communication in caves and mines does not take place solely in the near field, but in the transition zone, when the analysis is more difficult. There are also several different antenna orientations to consider.

The equations that describe the field due to a magnetic dipole in or over a conducting medium were derived in the 1960s. One such solution requires the integration of the product of a complex variable and a Bessel function. In the 1960s such computation was both slow and expensive. However, with the powerful personal computer applications now available, desk-top iterations are entirely feasible, and we can investigate the field behaviour at a level of detail that was previously not possible.

We do not need to specifically consider how to describe the secondary fields because we would expect these to arise as a result of the model we choose to use, simply by applying Maxwell's equations and the appropriate boundary conditions. In some circumstances, it is convenient to separate the primary field – which we assume to be the field from a static dipole with time-variation added. This is the approach taken by (Wait, 1982), although (Burrows, 1978) adopts a slightly different procedure, defining a *principal field* as that due to the dipole over a perfectly conducting medium. This is a suitable approach for calculating the above-ground far-fields, but is clearly not helpful for submerged dipoles.

1.4.5 Bibliography

The behaviour of static and slowly-varying fields was studied by James Clark Maxwell, who analysed conductivity in three dimensions, and in stratified media, in his treatise originally published in 1873; (Maxwell, 1954) Vol.1, Part 2, §7–§9. LF radio broadcasts in the first decade of the 20th century soon demonstrated that propagation was affected by the orientation of the transmitting antenna and the presence of a conducting ground plane. Experiments on radio reception in mines were undertaken prior to the first World War (Gibson, 1996b) although, as noted above, it was not until the 1960s that mine communications systems became practical. The theoretical problem was considered by (Sommerfeld, 1909; 1926)² who derived general expressions for the fields from infinitesimal dipoles on the surface of a semi-infinite flat homogenous earth. Sommerfeld derived an integral expression that bears his name, and many subsequent researchers have drawn on his work.

There are several textbooks – all out of print now – that cover the topic of VLF and LF propagation, although the emphasis is usually on the surface fields.

Dipole Radiation in the Presence of a Conducting Half-Space (Baños, 1966) is a detailed monograph, giving a mathematical treatment of the problem under all conditions of

2 Given as references in (Wait, 1969) and (Burrows, 1978).

conductivity and distance. There is more detail than we need for our present studies, for which Wait's formulations (as will be discussed below) are sufficient. As the title suggests, Baños does not focus on the problem of transition-zone fields below the ground, referring only briefly to this problem, in his last chapter.

VLF Radio Engineering (Watt, 1967) provides information for engineers to build and operate VLF broadcast systems. The sections on transmitter design are therefore not relevant to us, but those on receivers and atmospheric noise contain much useful practical information.

Unlike some more elementary geophysics texts, which only consider the quasi-static operation of induction loops, **Electrical Methods in Geophysical Prospecting** (Keller and Frischknecht, 1966) §6:38 gives a full treatment of the problem, discussing the primary field from an induction loop and the secondary field due to induced eddy currents. The purpose of the analysis is geophysical prospecting, rather than communications, but the principles are similar. Expressions are derived for the mutual coupling between loops of different orientations, placed above the surface of the ground; both in the general case (where numeric evaluation is required) and for some special cases (e.g. coplanar loops lying on the surface of the ground) where a closed form of the equations can be obtained.

ELF Communications Antennas (Burrows, 1978) describes the design of transmitting and receiving antennas for far-field communication in the range 30–300Hz. Burrows begins with a field analysis, which distinguishes between a 'principal' field, which would be radiated by an antenna in proximity to a perfect reflector, and a secondary field that results from the less-than-perfect conductivity of the ground. Burrows' primary consideration is the atmospheric propagation of the far-field, and his formulas are approximations, valid for distances much larger than the skin depth. At ELF, the antennas are much larger than we would consider for a short-range LF or VLF system, involving airborne electric dipoles or large ground arrays. One example of which is the antenna array for project *Sanguine* – a system proposed in the 1960s for submarine communications at ELF, consisting of a 100km square array of end-grounded electric dipoles carrying a distributed power of 10MW. (Llanwyn Jones, 1985). As with (Watt, 1967) the sections on receiver antenna design are more relevant to our situation, although Burrows discusses air-cored induction loops and ferrite-cored solenoids only briefly.

The person most associated with the analysis of electromagnetic sources in proximity to a conducting medium is Prof. J.R. Wait. In **Geo-Electromagnetism** (Wait, 1982) §3, Wait presents the topic in a similar way to (Keller and Frischknecht, 1966), although with a greater emphasis on the study of the electromagnetic fields than on the geophysical characteristics of the earth. Of particular interest is Wait's summary of the coupling between loops in different configurations. However, in this book, Wait considers only loops placed above the surface of

the ground. For an analysis of the above-surface field due to a horizontal underground transmitter (or vertical magnetic dipole, VMD), we need to refer to (Wait, 1969; 1971), and for the field due to a buried vertical loop (or horizontal magnetic dipole, HMD), to (Wait, 1972).

Of the above titles, several (Baños, 1966; Keller and Frischknecht, 1966; Watt, 1967) are works in Pergamon Press's *International Series in Electromagnetic Waves*. A fourth book in the series, *Electromagnetic Waves in Stratified Media* (Wait, 1962) studies the propagation of VLF waves in media whose properties vary in one direction. Although horizontal stratification in rock is mentioned in the introduction (and has been studied by Wait elsewhere) the book is mostly concerned with the stratification of the earth's atmosphere, and how this affects VLF signals. (Burrows, 1978) is a title in the IEE's *Electromagnetic Waves Series*. One other work in that series is worth mentioning – *Leaky Feeders and Subsurface Radio Communication* (Delogne, 1982) – although, as the title suggests, it is concerned with cable, rather than wireless communications.

1.4.6 Is Sub-Surface 'Radio' a Radio Phenomenon?

We have seen that, if the transmission distance is well within the near-field, the mode of operation for loop antennas will be predominantly magnetic and the methods associated with magnetic induction can be applied to describe the behaviour. That is, we assume that the field from a loop antenna is that of a static magnet, but with added time-variation; and we apply Faraday's law of induction. Although we use the word 'radio' to loosely describe such 'wireless' communication, one might infer from the word 'radio' that radiation was the primary phenomenon involved rather than magnetic induction. Moreover, is it actually *wrong* to talk of 'sub-surface radio' and 'cave radio-location' as some people assert; or is induction just one manifestation of a general radio phenomenon?

Partly this is a matter of semantics, with phrases such as 'induction radio' only serving to cloud the issue still further. We could suggest that 'wireless' may be a less contentious description than 'radio' and, to support our case, we could draw on the fact that 'wireless' may be a contraction for 'wireless radio'. The *Chambers 21st Century Dictionary* (Robinson, 1996) gives the standard definition of radio waves as "electromagnetic waves with a wavelength of 1mm to 10km". This corresponds, in air, to frequencies upwards of 30kHz. In a good conductor, the wavelength is 2π times the skin depth δ ; and skin depth is usually in the range of a few metres to a few hundred metres for sub-surface signalling. So, from this definition, electromagnetic communications in rock would be classed as radio.

An objection might be made, that 'radio' does necessarily imply electromagnetic radiation. However, the 17th-century word radiation predates electromagnetic phenomenon (e.g. beauty

can be radiated), and comes from the Latin *radiare* – to shine – and, ultimately, from *radius* – a spoke or ray – so the concept of a wave (originally a ‘Hertzian’ wave) as emanating from a source may have more to do with a simplistic concept of a ray than a particular notion of energy propagation.

Moreover, it is important to note that magnetic induction does not fully describe the phenomenon – for example, it does not explain how radio waves might ever be generated nor, in a conducting medium, does it easily demonstrate eddy current effects; and, starting with this quasi-static model we have no justification for simply combining it with a skin-depth attenuation term.

We have stated that sub-surface communications frequently take place in the transition zone, where the distance is too far for the system to be described purely as ‘quasi-static’. It will be shown later that, in a uniform conductive medium, there is an optimum distance for communication of around 2.8 skin depths (or 0.45λ) which falls within the transition zone from near to far field. At this order of distance the inverse-cube law of field strength v. distance begins to break down and, although energy radiation may not be significant, this effect of optimum frequency is not evident from a simple quasi-static model.

The radiation from a small loop is usually very low. For example, a loop with a specific aperture³ of $10m^2/\sqrt{\Omega}$ will radiate just $1\mu W$ for each $25W$ that is dissipated by ohmic losses in the antenna wire. If this antenna were submerged in a radome of diameter $10m$ in ground with a conductivity of $0.01S/m$ then approximately 0.3% of the power ($75mW$) would be dissipated by eddy current losses in the medium⁴. We can see that both the radiation resistance and this external resistance, which represents the eddy current losses, are but a small fraction of the ohmic resistance that we use to gauge the near-field strength. It seems that the eddy-current losses, being the far larger effect, could be responsible for the breakdown of the quasi-static model in the transition zone. However, we can only calculate these losses by using a complete field expression that allows for electromagnetic radiation.

The physicist and caver, Harold Lord, who did early radio experiments in the Gouffre Berger (Lord, 1969), was firmly of the view that subsurface induction communication was not a radio phenomenon. But Lord’s experiments were at audio frequencies where transition-zone effects would not have been apparent. Lord also insisted that multi-turn loops did not radiate, and used this to back up his assertion. But his argument, that the turns “destroy the phase and spatial relationship between the magnetic and electrical components of the generated fields”

³ Specific aperture is defined in chapter 3. A loop of diameter $1m$, made from $250g$ of copper will have an specific aperture of $10m^2/\sqrt{\Omega}$. (The number of turns does not affect the specific aperture).

⁴ See §3.4.3.

(Lord, 1988) is specious, and does not apply to ‘small’ loops where the wire length is much less than a wavelength. The radiation resistance for a small multi-turn loop is quoted, for example, in (Kraus, 1950) §6.8 as simply N^2 times that of a single-turn loop.

In conclusion, we consider that unless a ‘wireless’ system is truly quasi-static ($r \ll \lambda$ and $r \ll \delta$) its analysis *does* depend on an understanding of electromagnetic waves and so it can truly be described as a ‘radio’ system.

It may be worth pointing out that the word ‘propagation’ could also be subject to scrutiny since its use could infer a transmission of energy. In isolation, near-fields do not ‘propagate’ in this sense, but in a conducting medium, energy is transferred to the medium in the form of induced eddy currents. A further discussion of this point would be pedantic.

1.5 Example of a Communications Link

The nature of the sub-surface medium – in particular its conductivity – can vary over more than three decades in value. Similarly, the range of frequencies we may wish to use covers nearly three decades. A portable beacon or pager will use a small antenna, but a mine communications system may utilise a large, fixed antenna; both the antenna size and the power requirements could also be expected to range over three decades. It will therefore be appreciated that the example of a communications link, given below, does not cover all situations.

1.5.1 Transmitter Antenna Design

A portable induction loop of about 1m in diameter and weighing 0.2kg will, if there are no losses, produce a magnetic moment of 30Am^2 for a power dissipation of 10W. However, a larger, fixed installation might use a loop of diameter 50m, weighing 8.5kg, and producing a moment of $30,000\text{Am}^2$ for a power dissipation of 100W and resulting in a range ten times greater, if the field were to obey a pure inverse cube relationship⁵.

For applications such as portable beacons we may decide to use a small ferrite rod instead of an air-cored induction loop. A ferrite rod of length 200mm and diameter 24mm will typically produce a magnetic moment of 30Am^2 for a power dissipation in the winding of some 20W. However, at 30kHz, the eddy current losses in the core would be of a similar magnitude. If the same mass of ferrite were used in a rod of diameter 8mm and length 1.8m then the magnetic moment – which is limited by the saturation flux density – would remain unaltered but, because

⁵ Clearly the magnetic moment also depends on the number of turns of wire and its resistance, but we will show, in chapter 3, that these parameters are superfluous when we consider the mass, radius and power dissipation of the loop, as we have done here.

the core shape is more favourable, the power consumption to achieve this moment would be only 90mW, excluding the core losses.

Such a low power transmitter might be necessary in some circumstances but, to avoid significant core losses, operation may be limited to frequencies below 1kHz and so the antenna would be suitable only for data and beacon use, rather than speech communications. Alternatively, a very low loss antenna material could be used, such as amorphous metal tape, although this is extremely expensive.

1.5.2 Receiver Antenna

Similar arguments apply to the design of the receiver antenna. In an ideal design – as with all communications systems – the receiver antenna should be large enough so that the received signal swamps the locally generated noise due to the antenna resistance and the amplifier. In this situation the dominant contribution to the signal/noise ratio is the external or atmospheric noise. Therefore, as a rule, the maximum useful antenna size is related to the level of atmospheric noise.

Atmospheric noise is often described by its atmospheric noise temperature ratio. In chapter 6 we will derive a relationship between F_a and the ‘maximum useful size’ of the antenna, which we describe using the parameter of specific aperture, which we will define in chapter 3. One of the salient points of chapter 6 is that, for the expected range of F_a , the maximum useful size of a receiver antenna on the surface of the earth is quite small. For example, at 100kHz at a ‘quiet rural’ receiving location, if the external noise was as low as 82dB [dB above kT_0], this would indicate that our receiver antenna could be about the size of the 1m loop in the example above, but need be no larger. If the noise level was 20dB higher, the antenna could be 20dB (ten times) smaller (in terms of specific aperture). However, on a quiet day, this small antenna would have a limited range.

For a sub-surface receiver, a larger antenna can be used. If we consider that the external noise will be attenuated at 8.7dB per skin depth then the maximum useful size of the receiver (expressed in terms of specific aperture) will increase at a similar rate.

Unfortunately, as we will discuss in chapter 6, we cannot make use of the standard tabulated noise data that was collected using electric field antennas, and so we cannot draw any useful conclusions about noise until we have carried out a programme of measurements using magnetic field antennas.

The concept of a maximum useful size of the antenna arises out of a comparison between the atmospheric noise and the thermal noise of the antenna. In addition we must aim for the amplifier noise to be insignificant relative to the other noise sources. In chapter 6 we will

introduce the concept of ***noise-matching*** where it is shown that, for minimum noise, the antenna should present a source resistance that is similar to the ratio of the amplifier's noise voltage to noise current spectral density.

1.5.3 Propagation Losses

We stated earlier that the skin depth attenuation amounted to 8.7dB per skin depth, but this only applies to a plane electromagnetic wave. In chapter 2 we will show the attenuation of the field from an induction loop antenna includes three terms:

- i) the inverse cube law attenuation,
- ii) the exponential skin depth attenuation,
- iii) a complicated function of skin depth which *reduces* the effect of (ii)

For example, at 2.5 skin depths, a plane wave – for example, the atmospheric noise field – would be attenuated by 21.7dB due to skin depth effects. However, the coaxial field of an induction loop transmitter, over the same distance in a conducting medium, would only undergo 9dB of attenuation due to skin depth effects. (A different, and more favourable factor applies to a co-planar configuration). In addition, we would need to consider the inverse cube law attenuation.

A detailed discussion of skin depth will be given in chapter 2; and quantitative examples, demonstrating the transmitter power and the noise contributions to the received signal will be given in chapter 6. However, we can summarise by demonstrating a typical system, as shown in **Figure 1-2**, below.

1.5.3.1 Up-link operation

In this example, we assume a frequency of 20kHz and a ground conductivity of 1.25mS/m, which results in a skin depth of 100m. The atmospheric noise temperature ratio is $F_a = 148\text{dB}$, which results in an induced noise voltage in the above-ground receiver antenna of $123\mu\text{V}$. (This will be covered in more detail in chapter 6). The underground transmitter antenna consists of 34 turns of 7/0.2mm equipment wire on a diameter of 1m, and dissipates 66W of signal power in the winding. This results in a magnetic dipole moment of $m_d = 75\text{Am}^2$, and so the received signal, at the surface, will also be $123\mu\text{V}$. The signal to noise ratio (SNR) is therefore 0dB.

We can see, from **Figure 1-2**, that if the transmitter distance is halved, we gain 18dB of signal due to the lower cube-law attenuation, and a further 6.4dB due to the lower skin-depth attenuation, giving us an SNR of 24.4dB.

1.5.3.2 Down-link operation

For a down-link transmission, the atmospheric noise undergoes a 21.7dB attenuation due to the skin effect, so the SNR at 250m rises from zero to 21.7dB. At 125m we gain 24.4dB of signal, as above, but, additionally, the atmospheric noise is only attenuated by 10.9dB, so the SNR is 35.3dB (rounded to nearest 0.1dB).

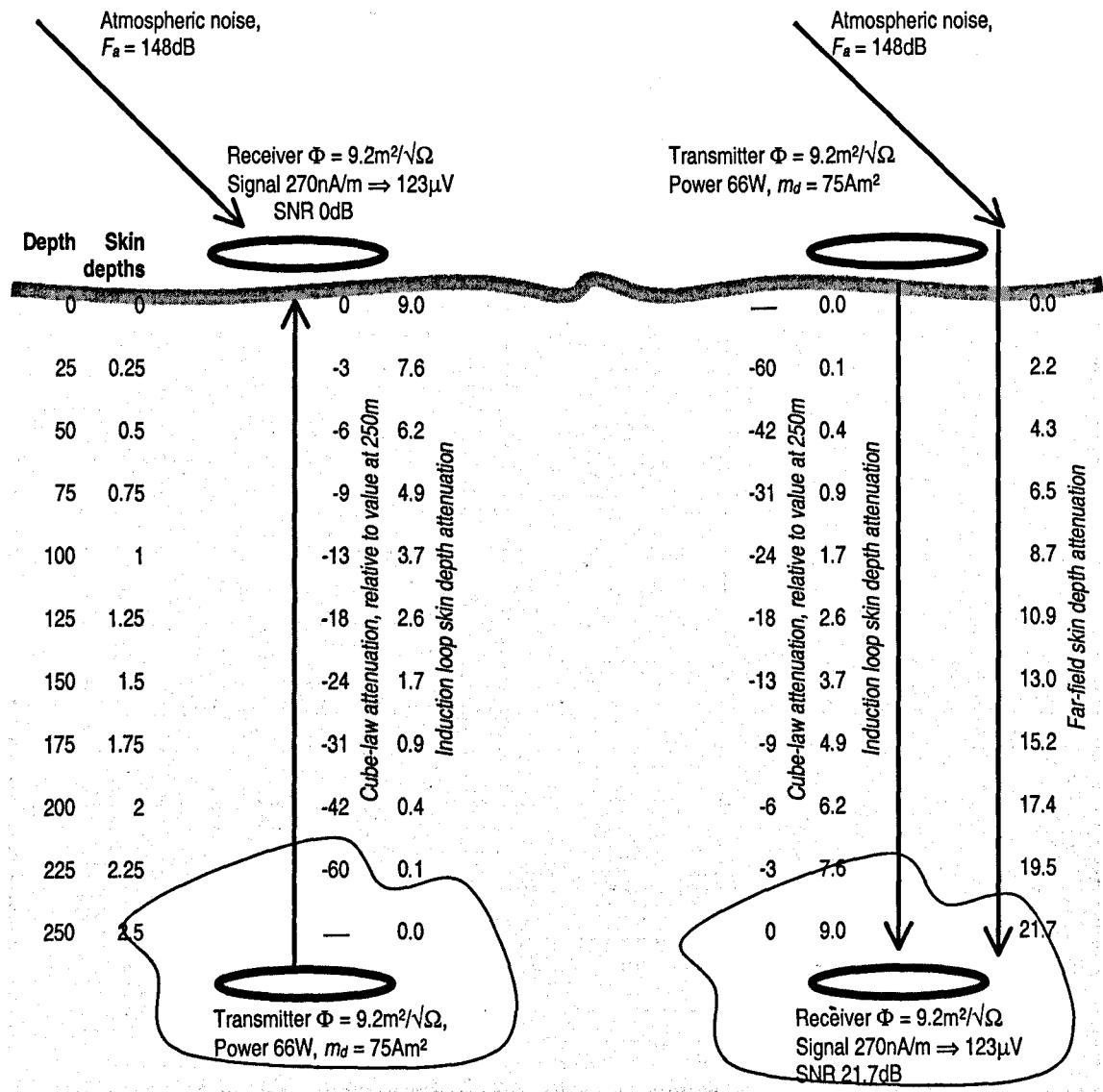


Figure 1-2 : A comparison of up-link and down-link SNR.

With the data given in the text, the up-link SNR is 0dB at 250m, whereas the down-link SNR is 21.7dB. At 125m the up-link improves to 24.4dB compared with 35.3dB for the down-link.

1.5.4 Optimum Frequency

For a communications system that uses induction loop antennas, and which operates mainly by magnetic induction, we know (from Faraday), that the induced signal will be less at low

frequencies. However, we also know that there will be increased signal attenuation at higher frequencies due to the skin effect. We can infer that there should be an optimum frequency, and therefore a *band-pass window*, where path attenuation is minimised. This will be discussed in detail in chapter 2, but we can demonstrate the effect by means of an example.

We shall consider a uniform medium with conductivity 1mS/m. At 100kHz the skin depth will be about 50m. Let us operate at 2.83 skin depths, so that the attenuation of the induction loop field will be 11dB. (The formula that expresses this will be derived in chapter 2). Now let us consider what happens if we alter the frequency. The effect is summarised in *Table 1-1*, below, where we can see that the increased skin depth attenuation is offset by the increase in induced signal. The effect is that of a broad, band-pass window centred at 100kHz.

The centre of the window is at the frequency for which the operating distance is 2.83 skin depths – a result that will be derived in chapter 2.

Frequency	25kHz	50kHz	100kHz	200kHz	400kHz
Skin depth	100.7m	71.2m	50.3m	35.6m	25.2m
Operating distance in skin depths	1.4	2.0	2.8	4.0	5.7
Skin depth attenuation	3.3dB	6.2 dB	11.0 dB	18.6 dB	30.3 dB
Induced signal (relative to 100kHz)	-12.0 dB	-6.0 dB	0.0 dB	6.0 dB	12.0 dB
Performance	-15.4 dB	-12.3 dB	-11.0 dB	-12.6 dB	-18.3 dB

Table 1-1 : Demonstration of optimum frequency effect.

The 'performance' is the induced signal minus the skin depth attenuation.

A conductivity of 1mS/m is assumed, and a fixed distance of 141m, i.e. 2.83 skin depths at 100kHz.

1.6 Chapter Summary and Original Contributions

A review of the propagation of electromagnetic fields in a conducting medium is an essential prerequisite to these studies, and this will be undertaken in *Chapter 2*. As a starting point we will derive the expressions for the electric and magnetic fields from a small magnetic dipole in a vacuum. We will then apply a simple modification to describe the fields in an infinite and uniform conducting medium, and will demonstrate some features of this model, such as skin depth and optimum frequency. The earth is better modelled as an infinite half-plane, for which we will use the results of J.R. Wait and others. A more complex analysis - allowing for stratified ground, and other generalities – will be avoided. Although the analysis is

a standard one, the fields can only be obtained by a lengthy numerical integration. ***Original work:*** Some originality can be claimed for the detailed computer simulations, which allow us to investigate the field behaviour at a level that was not possible when the equations were first formulated in the 1960s. We have made several original observations of the field characteristics as a result of these simulations, including an evaluation of the concept of an ‘optimum frequency’.

For near-field communications, such as we are studying here, the form of the fields depends on the antenna used to generate them. Before discussing antennas in detail it is convenient to define a ‘figure of merit’ by which we can assess and compare them, which we define in ***Chapter 3***. This concept – which we name specific aperture – was created for use with induction loop antennas, but we show that it is also useful in the context of radiating antennas, and that it is more appropriate than the existing term ‘effective height’ to describe a receiving antenna. ***Original work:*** specific aperture is an original concept and has proved invaluable in the description of transmitter and receiver antennas.

In ***Chapter 4*** we discuss some possible antennas. For most applications the choice is between an air-cored loop and a ferrite-cored solenoid, but other forms of antennas – electric dipoles and current dipoles are also possible. We consider briefly the use of toroidal antennas and, for radiolocation, the use of a rotating magnet. ***Original work:*** The comparison of antenna types, in this chapter and the two following, is presented within a structured and analytical framework. The discussions of toroidal and rotating-magnetic antennas demonstrate novel antenna systems that could be considered for experimentation.

In ***Chapter 5*** we discuss the transmitter design with particular reference to the power amplifier used to drive the antenna. The emphasis, in this thesis, is on small, portable induction loop antennas. Conventional speech communication usually uses a tuned loop, but for wideband systems an untuned induction loop is required. We discuss the difficulty of driving an untuned loop, and we introduce a figure of merit that incorporates the bandwidth and equalisation of a tuned antenna. ***Original work:*** Although this analysis is straightforward, we do not believe that it has previously been presented in such a detailed or systematic way, that allows us to assess the performance of different antenna systems.

In ***Chapter 6*** we consider the receiver antenna and the effect of noise. We derive an expression for the noise factor of the antenna and we discuss how this is affected by our choice of antenna design. In particular we note the noise penalty of an untuned and inductive antenna and we discuss how noise-matching can be used to increase the signal to noise ratio. ***Original work:*** Although this analysis is straightforward, we do not believe that it has previously been presented in such a systematic way. In particular, the concept of noise-matching is developed

with reference to the problems of reactive source, which we do not believe has been attempted previously in such detail. The chapter concludes with some discussion of methods of mitigating noise and interference, including the original concept of utilising ‘quadrature phase space’.

In *Chapter 7* we introduce the subject of channel sounding, as an aid to investigating the propagation and background noise at different frequencies. We outline the design of a practical channel sounder with which it is intended to perform a detailed channel evaluation. *Original work:* The design of this equipment is entirely original work and includes some novel features such as the ‘ripple trap’ latches, implemented in a programmable logic device, that help to mitigate the problem of a slow ripple carry along a chain of counters. Another notable aspect of the design is the instrumentation amplifier that features a capacitance multiplier. This allows the gain-setting resistors to be lower, in order to reduce the circuit noise

The sounder makes use of a modified pseudo-random binary sequence. Cross-correlation with the inverse sequence at the receiver results in a system identification signal that is used to maintain synchronism with the transmitter. In *Chapter 8* we discuss the design of sequences used for system identification. *Original work:* The concept of an inverse sequence is not new, but we can claim originality for the simplified method of calculating the inverse, and for the signal/noise analysis that we present.

Some preliminary results, using the sounder simply as a spectrum analyser are described in *Chapter 9* but, clearly, there is much work still to do to in evaluating and characterising the communications channel.

In *Chapter 10* we present our conclusions, and recommendations for further work.

In addition to the original work mentioned above, we can claim originality for the overall presentation of the topic. Published studies of through-the-earth propagation have focussed either on theoretical analysis or on the reporting of practical measurements. In this thesis we have attempted to give a broader view, encompassing the system design, which will be of use to equipment designers, and we have tried to present the information in a structured and modular fashion.

1.7 References⁶

- Anon. (2001). Welsh News: Daren Cilau. *Caves & Caving* 90, 10.
- Anon. (2002). Heyphone Cave Rescue Communication System. Available at [caves.org.uk / radio / heyphone /](http://caves.org.uk/radio/heyphone/)
- Austin, B. A. (1973). Talking through Solid Rock. *Coal, Gold and Base Minerals of Southern Africa* 21(9), 31-27.
- (1987). Radio Communication through Rock. *Electronics & Wireless World* 93(1619), 943-946.
- Baños, A. (1966). *Dipole Radiation in the Presence of a Conducting Half-Space. [International Series in Electromagnetic Waves: 9]*. Oxford: Pergamon Press.
- Bedford, M. (2001). Introducing the Heyphone. *Caves & Caving* 91, 15-17.
- Bedford, M. and D. Gibson (1995). Talk-through Boxes. *CREGJ* 20, 4-6.
- Burrows, M. L. (1978). *ELF Communications Antennas. [IEE Electromagnetic Waves Series: 5]*. Stevenage: Peter Peregrinus.
- Delogne, P. (1982). *Leaky Feeders and Subsurface Radio Communication. [IEE Electromagnetic Waves Series: 14]*. Stevenage: Peter Peregrinus.
- Durkin, J. (1984). Electromagnetic Detection of Trapped Miners. *IEEE Communications Magazine* 22(2), 37-46.
- (1991). Earth Conductivity Estimates from through-the-Earth Electromagnetic Transmission Tests. *IEEE Transactions on Geoscience and Remote Sensing* 29(2), 300-307.
- ETSI (1994). *European Telecommunications Standard ETS 300 330*. Valbonne, France: ETSI Secretariat.
- Gibson, D. (1996a). How Accurate is Radio-Location? *Cave & Karst Science* 23(2), 77-80.
- (1996b). Radio Speleohistory. *CREGJ* 26, 3.
- (1999). Selected Further Reading: Cb Transverter. *CREGJ* 37, 23.
- Keller, G. V. and F. C. Frischknecht (1966). *Electrical Methods in Geophysical Prospecting. [International Series in Electromagnetic Waves: 10]*. Oxford: Pergamon Press.
- Kraus, J. D. (1950). *Antennas*. New York: McGraw Hill.
- Llanwyn Jones, S. (1985). Sending Signals to Submarines. *New Scientist*, 4 July 85, 37-41.
- Lord, H. (1969). Communications Underground: Inductive Systems. In *Manual of Caving Techniques*. ed. C. Cullingford. London: Routledge & Kegan Paul, pp. 195-233.
- (1988). Letter: Radio Communications through Rock. *Electronics & Wireless World* 94(1627), p458.
- Maxwell, J. C. (1954). *A Treatise on Electricity and Magnetism*. New York: Dover Publications. (Facsimile of Unabridged Third Edition, 1891, Oxford: Clarendon Press).
- Murphy, J. N. (1978). Underground Mine Communications. *Proc. IEEE* 66(1), 26-50.
- Naylor, G. (1999). The Nicola Mark II - a New Rescue Radio for France. *CREGJ* 38, 3-6.

⁶ References to *CREGJ* are to the *Cave Radio & Electronics Group Journal*; a non-peer-reviewed low-circulation magazine published by a special interest group of the *British Cave Research Association*, bcra.org.uk. The CREG journal is filed at the British Library as ISSN 1361-4800. *Cave & Karst Science* is the BCRA's peer-reviewed cave science journal, ISSN 1356-191X.

- (2000). Highest & Deepest 2000. *CREGJ* **42**, 3-4.
- Nessler, N. H. (1981). *Subsurface Position Finding with ELF Waves*. IEE Second International Conference on Antennas and Propagation, York: IEE.
- Newman, G. (1993). The Caves of Thunder. *Descent* **110**, 26-27.
- Pease, B. (1991). Measuring Ground Conductivity with a Cave Radio. *Speleonics* **4**(4), 4-6.
- RA (1988). *MPT 1337 Performance Specification*. London: Department of Trade & Industry, Radiocommunications Agency.
- Ristenbatt, M. P., E. Holland-Moritz and K. Metzger (1988). A New Post-Disaster Mine Communication System. *IEEE Transactions on Industry Applications* **24**(2), 204-211.
- Robinson, M., Ed. (1996). *Chambers 21st Century Dictionary*. Edinburgh: Chambers.
- Sommerfeld, A. (1909). Über Die Ausbreitung Der Wellen in Der Drahtlosen Telegraphie. *Ann. Physik* **28**, 665-737.
- (1926). Über Die Ausbreitung Der Wellen in Der Drahtlosen Telegraphie. *Ann. Physik* **81**(17), 1135-1153.
- Wait, J. R. (1962). *Electromagnetic Waves in Stratified Media*. [International Series in Electromagnetic Waves: 3]. Oxford: Pergamon Press.
- (1969). Electromagnetic Fields of Sources in Lossy Media. In *Antenna Theory - Part 2*. ed. R. E. Collin and F. J. Zucher. New York: McGraw-Hill, pp. 468-471.
- (1971). Electromagnetic Induction Technique for Locating a Buried Source. *IEEE Transactions on Geoscience and Electronics* **GE-9**, 95-98.
- (1972). Locating an Oscillating Magnetic Dipole in the Earth. *Electronics Letters* **8**(16), 404-406.
- (1982). *Geo-Electromagnetism*. New York: Academic Press.
- Watt, A. (1967). *VLF Radio Engineering*. [International Series in Electromagnetic Waves: 14]. Oxford: Pergamon Press.
- Williams, B. (1995). Lamb Leer Cavern 1880-90: The Lake and the Talking Machine. *Proceedings of the University of Bristol Speleological Society*. **20**(2), 135-151.

2 Propagation

Abstract: The magnetic field from an elementary magnetic dipole in a vacuum can be derived using the method of retarded potential. A modification to the field expression allows us to examine the field in an infinite, uniform conducting medium and to draw conclusions on antenna orientation and optimum frequency. To model sub-surface propagation more accurately, we then use a half-plane model to approximate the earth's surface. This requires that we must impose boundary conditions, resulting in a more complex expression for the field, which we evaluate numerically on a computer. Further observations are then made concerning the characteristics of the field and its application to the problem of sub-surface radiolocation.

In the previous chapter, we described a variety of underground communications systems – including cable-based and wireless systems. This thesis is mainly concerned with wireless systems using small, portable antennas and it is therefore necessary to analyse the manner in which electromagnetic fields penetrate a conducting medium, which is the subject of this chapter.

The communications systems we will be considering rely, for a large part, on near-field effects. In this situation, unlike that of conventional radio, the behaviour of the fields cannot be analysed without considering the type of antenna to be used. For the purposes of this discussion we will restrict ourselves to considering the magnetic field from a loop antenna, since this is the most suitable antenna for small, portable, communications systems. We will return to the subject of antenna systems in a later chapter.

2.1 Electric v. Magnetic Fields

A magnetostatic source will generate a magnetic field that will penetrate a perfect electrical conductor, whereas an electrostatic source will generate an electric field that is terminated by charges on the surface of the conductor, with *no* penetration of the conductor by the field. Extending this observation to less-than-perfect conductors and low frequencies, we can appreciate that although an electric field will penetrate a conductor to some extent, there will be an attenuation across the boundary. If ϵ is the real part of the permittivity of the medium and ϵ_0 is the permittivity of free space then, provided that the medium is a 'good' conductor, with $\sigma/\omega\epsilon \gg 1$, the electric field is attenuated by

$$\frac{\sigma}{\omega \epsilon_0} \quad (2-1)$$

as it crosses the boundary into the conducting medium¹ – e.g. see (Clemmow, 1973) §5.3.2. If the conductivity, σ , were 1mS/m and the frequency were 100kHz then the attenuation would be 45dB, which demonstrates that the electric field may not be a good choice for sub-surface communications. Of course, a more comprehensive analysis would have to take into account the properties of the transmitting and receiving antennas. This subject was debated in chapter 1. In the analysis which now follows, we will anticipate some of the discussion that will take place in chapter 4 (where we will discuss antennas) and will consider the magnetic field from a small loop antenna as being the most relevant to our needs.

2.2 A Simple Model of a Loop Antenna

Small portable loop antennas have traditionally been used for communications in caves, and larger structures have been used in mining applications. Although a loop antenna is necessary for radiolocation, it is by no means the only suitable antenna for general communications, as will be discussed in chapter 4. The loop antenna does, however, present a straightforward means to model an electromagnetic source – it does not interact with its surroundings by ‘stray’ inductance or capacitance and, provided it is small, it can be treated as a point source.

Appendix A2.1 gives a standard derivation of the electric (**E**) and magnetic (**H**) vector fields from a small magnetic dipole with magnetic moment m_d , in a vacuum, using the method of retarded potential. By ‘small’ we mean that the magnetic dipole can be considered to be a point source; that is, if the field point is at a distance r and the dipole is a current loop of radius a then we have $a \ll r$. In spherical polar co-ordinates (r, θ, ϕ) the fields at an angular frequency ω are

$$\mathbf{H} = \frac{m_d}{4\pi r^3} e^{-jk_0 r} \left\{ 2\cos\theta(1+jk_0 r)\hat{\mathbf{r}} + \sin\theta(1+jk_0 r - (k_0 r)^2)\hat{\theta} \right\} \quad (2-2)$$

$$\mathbf{E} = \frac{j\omega\mu_0 m_d}{4\pi r^2} \sin\theta(1+jk_0 r) e^{-jk_0 r} \hat{\phi} \quad (2-3)$$

where k_0 is the *wave number* given by ω / c or $2\pi / \lambda$. (c is the speed of light and λ is the wavelength).

We will begin our analysis by considering these fields, but we will later make modifications to the expressions to allow for a conducting medium. The magnitude and phase of (2-2) is

1 By ‘attenuation’ we mean the ratio of the field strength in the two media. The physical cause of the effect is that the electric field mostly undergoes *reflection* at the interface, with only a small proportion being transmitted across the boundary.

plotted in **Figure 2-1** for the two special cases of $\theta = 0$ (coaxial receiver) and $\theta = 90^\circ$ (coplanar receiver)². The x -axis for these graphs is the parameter $k_0 r$, denoted by T in the figure.

Close to the transmitter ($T \ll 1$) there is an inverse cube relationship between field strength and distance, and at large distances ($T \gg 1$) this becomes inverse linear for the coplanar field, and inverse-square for the coaxial field. For $T \gg 1$ the wavelength is established as 2π in the normalised scale of the x -axis, as we would expect. There is a curious phenomenon in the phase of the coplanar field, which establishes a phase lead that peaks at around $T = 1.5$ instead of the expected phase lag, and which continues to lead the coaxial field by 90° . This is associated with the establishment of a radiation field in the coplanar region, whereas the co-axial field for $T \gg 1$ is a residual induction field.

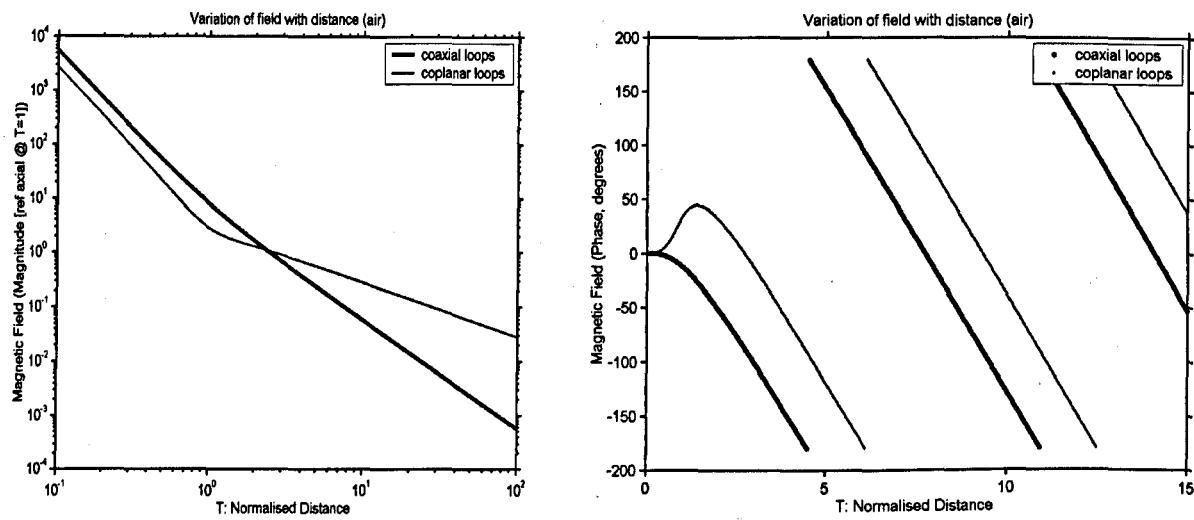


Figure 2-1 : Magnetic dipole field v. distance for a vacuum.
The co-axial and co-planar magnetic dipole field, magnitude (left) and phase (right).
The x-axis is the 'normalised' parameter $T = 2\pi r / \lambda$, where r is the distance from the transmitter.
The y-axis is normalised so that the co-axial field is unity at $T = 1$.

Equation (2-4) shows that the far-field magnitude along the axis of the dipole is zero. This is observable in **Figure 1-1** and **Figure 2-1** and must, in any case, be true from arguments of symmetry. This makes it doubtful whether the assertion in (Lord, 1988) is meaningful; that we could observe a change in orientation of the field along the axis of the dipole as we moved from the near to the far field. However, a change in field directions is observable at other positions. For example, at an angle of $\theta = \arctan\sqrt{2} \approx 55^\circ$ to the axis of the loop, the near-field is horizontal, at 90° to the axis, whereas the far-field is at $\theta + 90^\circ$ to the axis, as shown in **Figure 2-2**, below.

2 See **Figure 2-11** on page 45 for a representation of the coplanar and coaxial configurations.

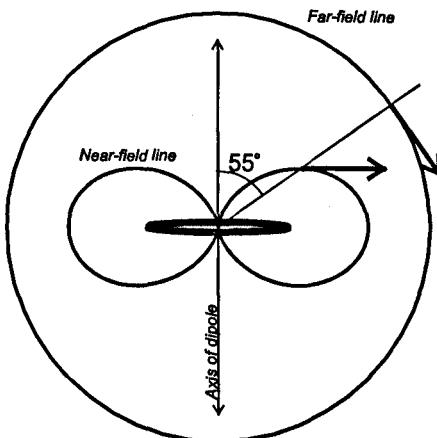


Figure 2-2 : Magnetic dipole, showing different direction of near and far-field lines.
The diagram depicts a vertical magnetic dipole. For a field point at 55° to the axis, the near-field is horizontal, whereas the far field line is at an angle of 55° to horizontal.

We shall show, later, that the presence of a conducting medium gives rise to additional field components, due to a secondary induced field. The field at the point depicted above then includes a non-zero vertical component with a different phase. This makes Lord's suggestion even less tenable, because any phase or polarisation effect due to the far-field would be masked by that due to the secondary induced field. This is important because the direction of the field line is used as an aid to location – for example, for tracing trapped miners or for surveying underground cavities for civil engineering purposes or for drawing cave surveys. We shall return to this topic later in the chapter.

2.2.1.1 Far or Radiation Field

Using an approximation of the form $(1 + x) e^{-x} \approx 1$ for $x \ll 1$, we can identify the two special cases of (2-2) that are observable in the above figures. With $k_0 r \gg 1$ the fields approximate to

$$\mathbf{H} = \frac{m_d k_0^2}{4\pi} \frac{e^{-jk_0 r}}{r} \sin \theta \hat{\theta}, \quad E_\phi = Z_0 H_\theta \quad (2-4)$$

with $Z_0 = \sqrt{(\mu_0 / \epsilon_0)}$. Here, the \mathbf{E} and \mathbf{H} fields are in phase and so convey energy.

2.2.1.2 Near or Induction Field

With $k_0 r \ll 1$ the fields approximate to

$$\mathbf{H} = \frac{m_d}{4\pi r^3} \left\{ 2 \cos \theta \hat{r} + \sin \theta \hat{\theta} \right\}, \quad \mathbf{E} = \frac{j\omega \mu_0 m_d}{4\pi r^2} \sin \theta \hat{\phi} \quad (2-5)$$

which, unsurprisingly, shows that the magnetic field obeys the Biot-Savart law for a current loop, and that the electric field follows the form of an electrostatic dipole. The \mathbf{E} and \mathbf{H} fields are in quadrature and so convey no power.

Although we invoked the concept of retarded potential, we could have derived the induction field expression directly from magnetostatic principles, which do not consider the finite speed of propagation. The induction field expression therefore seems to indicate that the propagation is instantaneous, and that phase of the field does not alter with distance. *Figure 2-3* shows that this approximation is only valid for a small region close to the transmitter. This suggests an alternative definition of the transition zone, based on phase shift, should we so desire it.

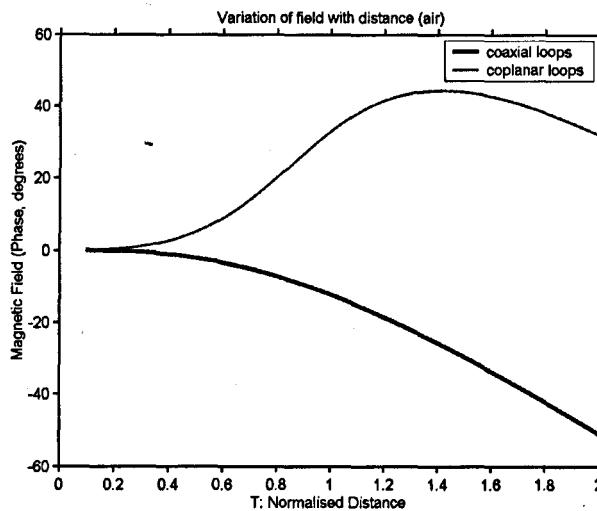


Figure 2-3 : Magnetic dipole field v. distance for a vacuum (expanded view).

2.2.1.3 Observations on the Inverse Cube Law

As we have seen, the magnetic near-field decays as the inverse cube of the distance. We can ask if there is any means of altering this rate of attenuation, to make it more favourable for us. For a wire carrying a steady current, the contribution to the magnetic field from a small element $I \delta l$ is found (from the Biot-Savart law) to be proportional to $1/r^2$ where r is the radial distance to the wire element. Integrating over a length of wire, this results in a field proportional to $1/r$. So, if we are close enough to the wire for it to look ‘infinitely’ long or, equivalently, for the return current path to be far more distant, then the field will decay as $1/r$. If we were far enough away for the wire to look like a small current element then the field would follow an inverse square law. However, this situation is not possible to achieve, because there must be a return path for the current and, if this is sufficiently close to the ‘outgoing’ current then the magnetic field will tend to cancel. A closer look at this situation reveals that it is not precisely the fact that there *is* a return current path, but that the resultant field has to be resolved as the sum of two vectors, that gives rise to the inverse cube law. For example if, instantaneously, one half of the circumference of a small current loop were to carry no current, then the axial field caused by the current in the other side would still resolve to give an inverse cube field along the axis of the loop. A brief discussion of this problem appeared in (Gibson, 1996a).

We can conclude that the advantage of using a very large induction loop is not only the size of the field that it generates but that the field decays less rapidly very close to the loop. Both of these characteristics are caused by the distant 'return path' of the current. However, we should note that this argument applies only to the near-field in air. At distances where the far-field predominates we saw, in *Figure 2-1* that the loop orientation can be changed from a coaxial to a coplanar configuration, where the field decays with an inverse linear law. And we will see, in §2.2.3, that this inverse-linear law is over-ridden by the effect of skin depth attenuation in a conducting medium.

The magnetic near-field from an electrostatic dipole decays as the inverse square of the distance; and, although we would not normally choose to use an electric dipole as a transmitter, we can see that there may be advantages in some situations. (See, for example, the discussion on toroidal antennas in §4.8). The success of the *J*-field antenna (§4.1.3), which has some of the attributes of both a large induction loop and an electric dipole, can also be attributed to the fact that the magnetic field obeys an inverse square law.

2.2.2 Propagation in a Conducting Medium

The field equations derived above are applicable in a vacuum or non-conducting medium. To modify the expressions for a conducting medium we can assign a complex value to the wave number, k , which we substitute into the field equations to allow us to describe the propagation in an infinite, uniform, conducting material. This process will now be described.

2.2.2.1 Plane Wave Propagation and Skin Depth

When a monochromatic plane wave travels in a conducting medium it obeys the Maxwell-Faraday and Maxwell-Ampère relations, namely

$$\nabla \wedge \mathbf{E} = -j\omega\mu\mathbf{H}, \quad \nabla \wedge \mathbf{H} = j\omega\mathbf{D} + \mathbf{J} \quad (2-6)$$

and in a simple material, characterised by constant values for ϵ , μ and σ we can write the latter

$$\begin{aligned} \nabla \wedge \mathbf{H} &= (j\omega\epsilon + \sigma) \mathbf{E} \\ &= j\omega(\epsilon - j\sigma/\omega) \mathbf{E} \end{aligned} \quad (2-7)$$

suggesting that a conductive medium can be represented by replacing the permittivity by a complex quantity defined by

$$\epsilon' = \epsilon - j\sigma/\omega \quad (2-8)$$

Substituting this complex ϵ' in the expression for the wave number (A2-2) results in a complex wave number

$$k' = \sqrt{\omega^2\mu\epsilon - j\omega\mu\sigma} \quad (2-9)$$

We can substitute this into the $\exp(-jk_0r)$ term in the field equations, but it is convenient to separate the real and imaginary parts first, to give

$$k' = k_r - j\gamma \Rightarrow e^{-jk'r} = e^{-jk_r r} e^{-\gamma r} \quad (2-10)$$

where k' is the (real) *wave number* and γ is the *attenuation coefficient*.

It is usual to consider the reciprocal of the attenuation coefficient as an indicator of the penetration of the field. $1/\gamma$ has the dimension of length, and it is referred to as the *skin depth* and we define it in terms of the properties of the medium, μ , σ and ϵ . At low frequencies, or in a good conductor, which we define by

$$\sigma/\epsilon\omega \gg 1 \quad (2-11)$$

we can approximate (2-9) to

$$k' \approx \sqrt{-j\omega\mu\sigma} \quad (2-12)$$

$$k' = \frac{1-j}{\delta}, \quad \text{with } \delta = \sqrt{\frac{2}{\omega\mu\sigma}} \quad (2-13)$$

This result can also be obtained by ignoring the displacement current term in (2-7) and using 2-6 to derive

$$\nabla^2 \mathbf{H} = j\omega\mu\sigma\mathbf{H} \quad (2-14)$$

which is in the form of the *diffusion equation*. Details are given in, for example, (Ferrari, 1975) §7.5.

Substituting k' from (2-13) into $\exp(-jk'r)$ in (2-4) demonstrates that the wave propagates with velocity $v = 2\pi\delta f$ and therefore with wavelength

$$\lambda_{rock} = 2\pi\delta \quad (2-15)$$

and exponential attenuation $\exp(-r/\delta)$. There is a definite physical interpretation that a plane wave is attenuated, in a good conductor, at a rate of $1/e$ or 8.7dB per skin depth; that the wavelength is much shorter than in free space, that skin depth depends on frequency and that the velocity is proportional to \sqrt{f} . However, it will be shown later that this physical interpretation of the figure of merit that is skin depth must be applied with caution to a near-field situation.

At high frequencies, or in a poor conductor, such that $\sigma/\omega\epsilon \ll 1$ we can make another convenient approximation to k' , as listed in the table below. The condition

$$\sigma/\epsilon\omega_c = 1 \quad (2-16)$$

is sometimes used to define a *characteristic frequency* ω_c for the medium. For frequencies much lower than this, the medium behaves like a 'good' conductor and displacement current is

negligible; whereas, for frequencies much higher than this, the medium behaves like a ‘poor’ conductor and displacement current is significant. The effect on skin depth is shown in *Table 2-1*, below. For frequencies where neither of the approximations is applicable, we can derive an exact expression for k' . The detail is shown in appendix A2.2.

Medium	Conditions	Skin Depth	Wavelength
In a “good” conductor,	with $\sigma/\epsilon\omega \gg 1$ or $\omega \ll \omega_c$	$\delta = \sqrt{2/\omega\mu\sigma}$	and $\lambda = 2\pi\delta$
In a “poor” conductor,	with $\sigma/\epsilon\omega \ll 1$ or $\omega \gg \omega_c$	$\delta = \sqrt{\epsilon/\mu} \cdot 2/\sigma$	and $\lambda = \lambda_0$
In free space,	with $\sigma = 0$	$\delta = \infty$	and $\lambda = \lambda_0$

Table 2-1 : Definition of skin depth in good and poor conductors.

2.2.2.2 Electrical and Magnetic Properties of the Ground

Conductivity: Electrical conduction through the ground takes place in three ways, by electronic (ohmic), electrolytic and dielectric methods; see, for example, (Telford, Geldart, Sheriff and Keys, 1976) §5. Most rocks, in their dry state, would be poor conductors, but they are usually porous, containing water and dissolved electrolytes. The resulting conductivity is provided by an actual transport of material (ions) and can be expected to be highly dependent on frequency.

Igneous rocks tend to have the lowest conductivity and sedimentary rocks the highest. The range of values is highly dependent on the characteristics of the rock and the water content. Most caves occur in limestone, which, in its dry state, has a conductivity of typically 1mS/m. Alluvium, sands and clays have a range from 1mS/m to 1S/m depending on water content. For communications in mines or tunnels we could expect a wider variation – sandstone and granite can vary from 1mS/m down to 1μS/m and lower if the material is dry.

In a bedded and jointed rock such as limestone we would expect the conductivity to be anisotropic, and for the rock to be more conductive along a bedding plane than at right angles to it. Geophysicists refer to the *anisotropy coefficient*, which is the ratio of maximum to minimum conductivity as a function of direction. For limestones and shales this can reach 1.2, (Telford *et al.*, 1976) §5.2.2.

Dielectric conductivity is the term used by geophysicists to refer to the effect of electrical permittivity. In traditional electromagnetic theory a displacement current flows in a non-conductor when an external electric field varies with time, and the resulting electrical polarisation can be electronic, ionic or molecular. This does not result in energy dissipation, and so does not affect our measure of conductivity. However, if the permittivity is significant, it may cause the characteristic frequency (2-16) to fall to a point where rock can no longer be considered a good conductor.

Permittivity: The relative permittivity (ϵ_r), or dielectric constant of rocks can vary widely and is, like conductivity, highly dependent on the water content, since water has $\epsilon_r \approx 80$. In many materials it is inversely proportional to frequency. Dry to moist clay has a typical range of $\epsilon_r = 7$ to 43 according to a table in (Telford *et al.*, 1976) §5.4.2. Geophysicists usually measure resistance at d.c., but measurements of dielectric constant may be made at around 100kHz.

Since the conductivity of wet rock can be significantly higher than 80 times that of dry rock it seems likely that the lowest values of characteristic frequency will be found in dry rocks. Taking a somewhat arbitrary description of a dry rock as $\sigma = 1\text{mS/m}$ and $\epsilon_r = 7$ we obtain a characteristic frequency of 2.6MHz, and we can deduce that the rock will be a 'good' conductor up to perhaps 260kHz, where the skin depth would be 31m. At higher frequencies – above, say, 26MHz – where the earth is a poor conductor, the skin depth would be 14m and independent of frequency. At 27MHz we would therefore expect CB radio to penetrate a short distance, which indeed, it has been observed to do.

It is apparent, from a study of the literature, that the conductivity and permeability of rock varies widely depending on the age, location and local conditions. We do not know of a survey of rock types that is sufficiently detailed to predict the conductivity of regional cave-bearing limestones. One possibility for further work would be to undertake such a field survey, and to include mining regions, and areas where railway tunnels exist in rocks other than limestone.

Permeability: The magnetic permeability of the ground varies very little from its free-space value unless magnetic ore bodies are present. It is usually sufficient to assume that it is μ_0 . A typical range of skin depths that we might encounter is shown in the graph and table below.

Conductivity (mS/m)	Frequency (kHz)	Skin Depth (m)
100		23
10	5	71
1		225
0.1		712
100		8
10	40	25
1		80
0.1		252
100		4
10	200	11
1		36
0.1		>113 *

Table 2-2 : Skin depth, as a function of conductivity and frequency.

* Note that at 200kHz and 0.1mS/m we are within a decade of the characteristic frequency and so the rock can no longer be considered to be a good conductor.

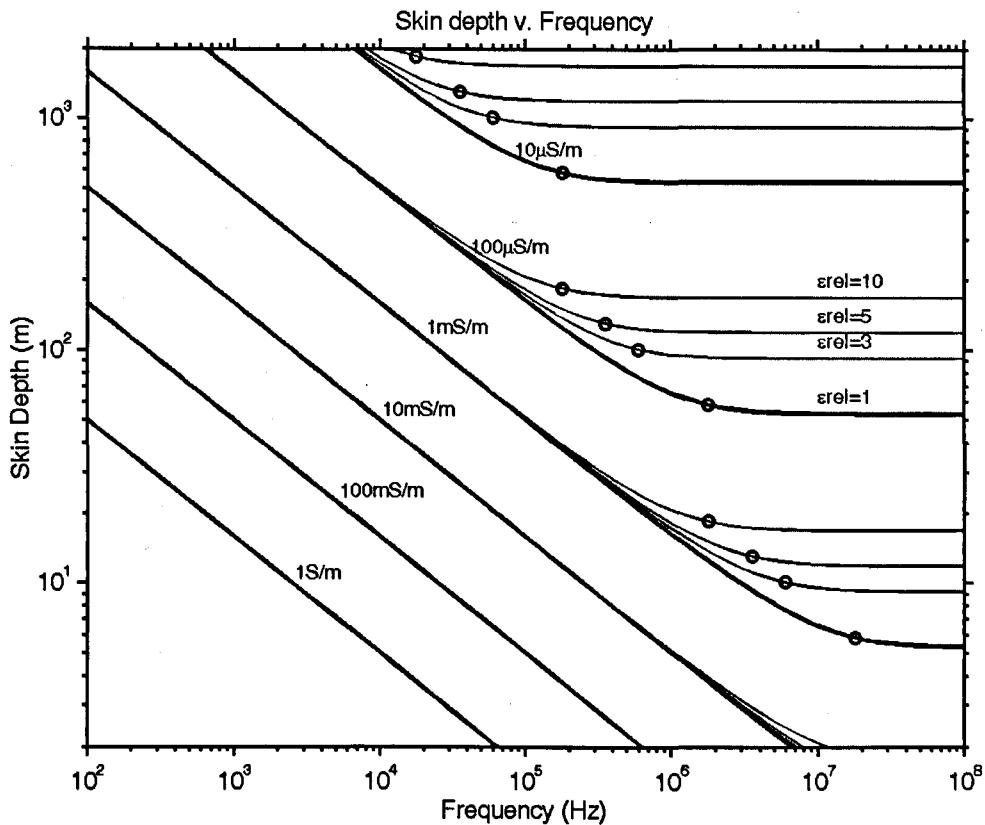


Figure 2-4 : Skin depth, as a function of conductivity and frequency.
At 100kHz and 1mS/m the skin depth is 50m. Above the 'characteristic frequency' – marked \circ – the medium is no longer considered to be a 'good' conductor. About a decade below this frequency the skin depth departs from its 'good conductor' value and starts to become dependent on permittivity rather than frequency.

The graph is based on the full expression for skin depth given in appendix A2.3.

2.2.3 Fields from a Magnetic Dipole in an Infinite Conducting Medium

We can assume that, in an infinite conducting medium, we do not have to consider boundary conditions, and so we can substitute a complex wave number (2-13) directly into the field expressions (2-2) to arrive at an expression for the field. This procedure does not address the nature of medium in the immediate vicinity of the source, which must presumably be an insulator if the antenna is to function as an isolated magnetic dipole. We will introduce the dimensionless parameter

$$T = \frac{r}{\delta} \quad (2-17)$$

to obtain

$$\mathbf{H} = \frac{m_d}{4\pi r^3} e^{-jT} e^{-T} \left\{ 2\cos\theta(1 + (1+j)T)\hat{\mathbf{r}} + \sin\theta(1 + (1+j)T + 2jT^2)\hat{\theta} \right\} \quad (2-18)$$

$$\mathbf{E} = \frac{j\omega\mu_0 m_d}{4\pi r^2} \sin\theta(1 + (1+j)T)e^{-jT} e^{-T} \hat{\phi} \quad (2-19)$$

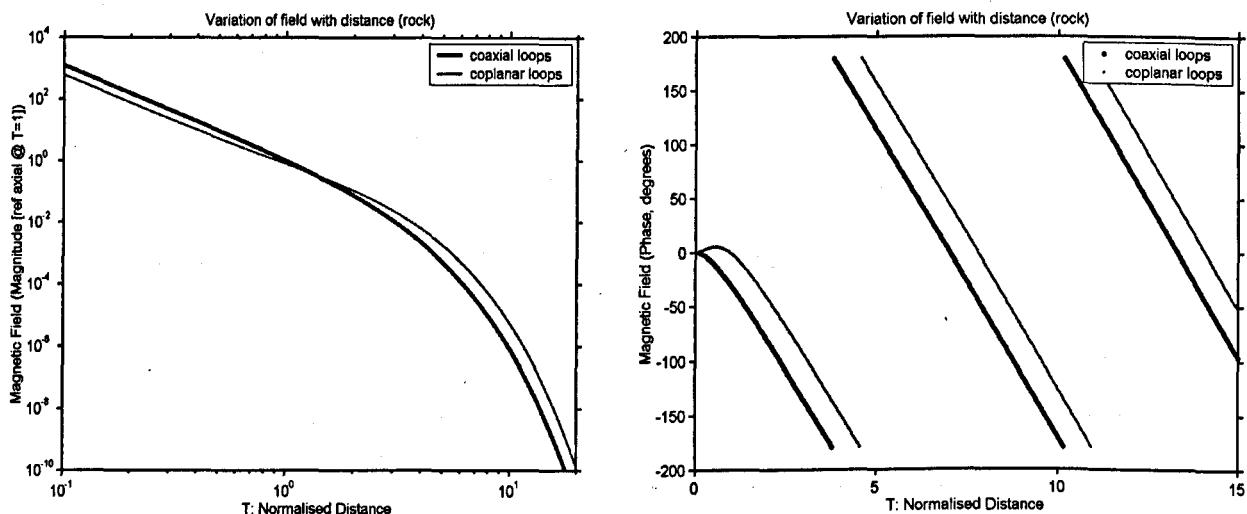
As before, we can apply near-field and far-field approximations, with $T \ll 1$ and $T \gg 1$. For the near field, the expressions are identical to (2-5) whilst, for the far field, we obtain

$$\mathbf{H} = j \frac{m_d}{2\pi r} \frac{1}{\delta^2} e^{-jT} e^{-T} \sin \theta \hat{\theta}, \quad E_\phi = \eta \left(\frac{1+j}{\sqrt{2}} \right) H_\theta \quad (2-20)$$

where η is the *surface impedance* of the medium, defined by

$$\eta = Z_0 \sqrt{\frac{\epsilon \omega}{\sigma}} \quad (2-21)$$

Equation (2-20) confirms common knowledge, that the far-field signals propagate with wave number $1/\delta$ (i.e. wavelength $2\pi\delta$), with an exponential attenuation described by the skin depth; that the electric and magnetic fields are 45° out of phase, and that the wave impedance³ is $\eta\sqrt{j}$. The near field approximation in air, of $k_0 r \ll 1$, becomes $T \ll 1$ (i.e. $r/\delta \ll 1$) in the conducting medium. Because the transition distance is defined as $\lambda/2\pi$ we can now see that the transition distance is exactly one skin depth, at $T = 1$.



*Figure 2-5 : Magnetic dipole field v. distance for a conducting medium.
The co-axial and co-planar magnetic dipole field, magnitude (left) and phase (right).
The x-axis is the 'normalised' parameter $T = r/\delta$, where r is the distance from the transmitter.
The y-axis is normalised so that the co-axial field is unity at $T = 1$.*

The magnitude and phase of (2-18) is plotted in *Figure 2-5* for the two special cases of $\theta = 0$ and $\theta = 90^\circ$, as before. As with the case of the dipole in a vacuum, we can observe that close to the transmitter ($T \ll 1$) there is an inverse cube relationship between field strength and distance. However, the situation for $T \gg 1$ is completely different, and demonstrates an exponential skin depth attenuation. As before, there is a slight phase lead in the coplanar field

3 We can also write the impedance as $(1+j)/\sigma\delta$, which is one of several equivalent forms that can be seen in the literature.

and, in the far field, it leads the coaxial field by 45° instead of 90° (although the coaxial field soon becomes insignificant in magnitude).

Figure 2-6 shows that there is only a very small region, close to the transmitter, where the phase lag is insignificant. It seems unreasonable to attribute the phase lead in this region to the establishment of a radiation field and so we surmise, without further investigation, that this represents the secondary induced field in the rock. As we will see in an example in §3.4, the dissipation due to eddy currents is far greater than the power conveyed in the radiation field.

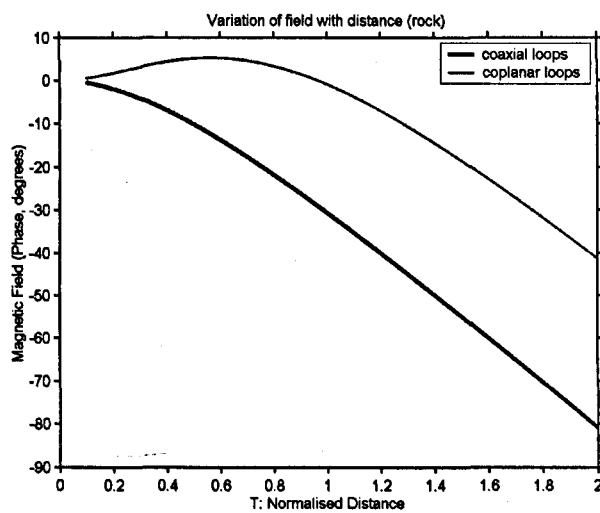


Figure 2-6 : Magnetic dipole field v. distance for a vacuum (expanded view).

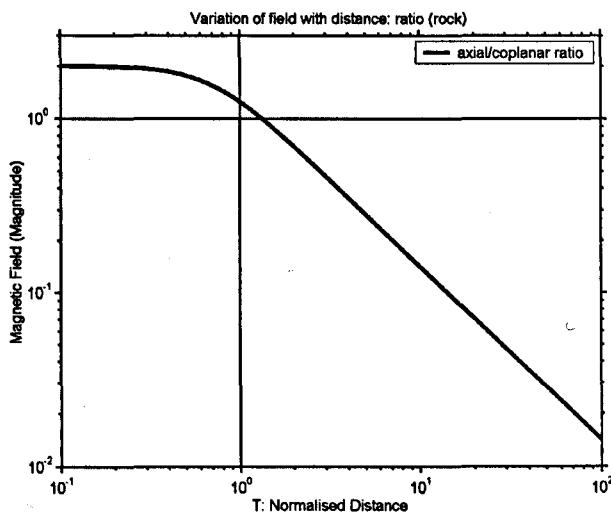


Figure 2-7 : Ratio of coaxial to coplanar field v. distance.
At distances above a skin depth ($T = 1$) there is an advantage in using a coplanar orientation because the co-axial field decays rapidly.

Figure 2-5 showed that the co-axial near-field was stronger, in air, than the co-planar field, but that outside this region the co-planar field was stronger. Although cave radio experimenters have known the practical significance of orientation in the transition zone for some time⁴ a justification was attempted (Drummond, 1987b) solely from a knowledge of the field characteristics in air. Here we have demonstrated analytically that the result holds when the field is in a conducting medium and modified by skin depth attenuation. *Figure 2-7* shows how the ratio of the coplanar to coaxial field varies with distance in a conducting medium.

2.2.3.1 Validity of Near Field Approximation

Inspecting the equations for the near-field (2-5) and far-field (2-20) seems to suggest that the near-field is unaffected by skin depth effects. Lest we believe the proposition that the rock somehow ‘knows’ whether the field passing through it is ‘near’ or ‘far’ in nature, we should note that the near-field expression does not include any phase information, suggesting that the field travels with infinite speed; and neither does it appear to take any account of transfer of energy by magnetic induction and eddy current effects. The explanation for these omissions is, of course, that the near- and far-field expressions are merely approximations. The near-field, or quasi-static approximation is only completely correct for the static situation at zero frequency; the term ‘quasi-static’ meaning that we just add a time variation⁵ to give us an approximation for a very low frequency or small distance ($T \ll 1$). This approximation starts to break down as soon as the phase lag, shown in *Figure 2-1* to *Figure 2-6*, departs significantly from zero.

We could deduce from (2-18) that if we wished to consider only the *magnitude* of the field and not its phase then, for $T \ll 1$, a reasonable approximation might be to multiply the quasi-static field by a skin depth term, $\exp(-T)$. *Figure 2-8* shows that this not a very good approximation, even at $T = 0.1$, and we must therefore be careful not to be misled into believing that it is a correct physical representation of the situation simply to add skin depth attenuation to the quasi-static expression. However we can note from (2-18) that a skin depth term *does* multiply the complete field expression. A more instructive ‘near-field’ approximation would be to assume not that $T \ll 1$, but that $T^2 \ll 1$. In this situation, the field approximation is that of a quasi-static field multiplied by

$$e^{-jT} e^{-T} (1 + (1+j)T) \quad (2-22)$$

⁴ It is also appreciated that a practical system may have to be oriented for minimum noise rather than maximum signal.

⁵ But we have not yet rigorously defined quasi-static, and there are several conflicting definitions in common use. See §2.2.3.2

which is a good representation of the co-axial near-field at all distances, and of the co-planar field at up to about a skin depth (*Figure 2-8*).

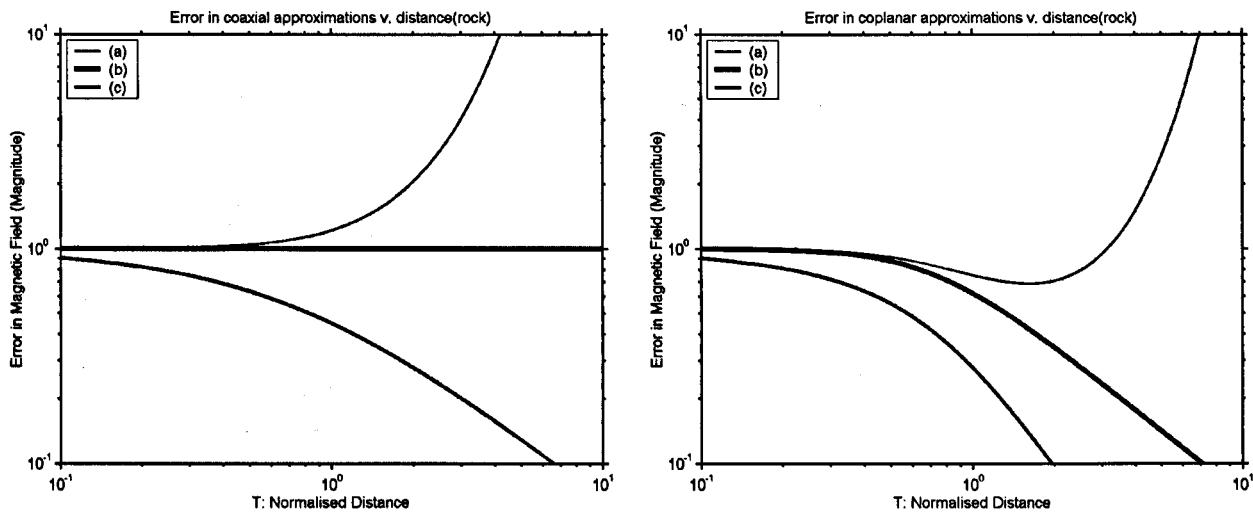


Figure 2-8 : Some approximations to the full-conducting-space model.

Left: co-axial field, Right: co-planar field. (a) quasi-static inverse cube law, (b) equation 2-22, (c) quasi-static field with skin-depth attenuation. In the far-field, the quasi-static expression (a) over-estimates the field strength, but the quasi-static + skin-depth approximation (c) under-estimates it.

The absence of the δ term in the quasi-static approximation has lead to some confusing definitions. In equations derived by (Wait, 1969; 1971) the occurrence of a term based on $j\omega\mu\sigma$ (see our definition of k' in 2-12) which Wait denotes by γ^2 leads to a convenient figure of merit of $1/\sqrt{(\omega\mu\sigma)}$ which, confusingly, is not the definition of skin depth that we are used to. Wait's earlier formulations use this to 'normalise' certain distance measurements and (Shope, 1991) also omits the factor of 2 from his version of (2-13)⁶. However, Wait later comments (Wait, 1982), p112, that "the choice of δ is somewhat arbitrary" and he re-inserts the factor of 2 to use the more 'conventional' definition of skin depth. The choice of δ may be arbitrary, but its usual definition is helpfully consistent.

2.2.3.2 Definition of Quasi-Static and other Field Approximations

We have described and named several field approximations, and it will be helpful to summarise their definitions. This is particularly important in the case of the *quasi-static approximation*, because various different conventions are in use.

Throughout this thesis, we use the term 'quasi-static' to mean that the electromagnetic fields follow the form of a static dipole (electrostatic or magnetostatic) with added time variation. As we have seen, this is only true at distances very close to the source. The traditional near-field

6 We will use the symbol T to distinguish Wait's normalisation from the definition we have adopted. See equation (2-29).

approximation, of $k_0 r \ll 1$ results in a quasi-static expression, but the near-field characteristics (e.g. lack of significant radiation) may be present over a wider region of space than that in which the quasi-static expression is valid.

The condition $k_0 r \ll 1$ is equivalent to $\omega r/c \ll 1$. The term quasi-static could be interpreted as meaning ‘slowly varying current’ but some authors appear to consider that this constraint on ω is satisfied, not by $\omega r/c \ll 1$, but by $\omega\epsilon/\sigma \ll 1$. However, as we have seen, this condition – and the approximation that results from it – does not serve to separate near-field from far-field effects. In a short letter on the subject, (Wait, 1995a; b) cautions against this undesirable convention, and suggests that the term ‘quasi-static’ should be used specifically for the condition $k_0 r \ll 1$. In this situation, at distances much lower than a free-space wavelength, the fields in air can be derived from a scalar potential, which fact Wait makes use of in his derivations. The complex part of the wave number can remain unrestricted, although the T term appears sufficiently often for us to consider the approximation $T \ll 1$.

Bannister⁷ used ‘quasi-static’ for $T \ll 1$, ‘near-field’ for $k_0 r \ll 1$ and ‘quasi-near’ for the region between the two; and Shope refers, misleadingly, to Wait’s formulations⁸ as being ‘quasi-static’, stating in (Shope, 1991) that

“The model discussed in this paper describes quasi-static induction fields; these are non-propagating fields. Because skin depth is usually associated with propagating waves it may seem inappropriate in this formulation...”

In this thesis, we will use the terms in the form listed in the tables that follow.

i) Distance of field point

Name of Approximation	Conditions	Circumstances when valid	See equation...
Quasi-static approximation	$T \ll 1$, $k_0 r \ll 1$	Fields are quasi-static in nature (i.e. static, with added time variation)	(2-5)
Near-field approximation	$T^2 \ll 1$, $(k_0 r)^2 \ll 1$	Fields show skin depth attenuation in a conducting medium, phase lag, and inverse square/cube laws	(2-22)
Transition distance	$T = 1$, $k_0 r = 1$	Arbitrary demarcation distance	
Far-field approximation	$T \gg 1$, $k_0 r \gg 1$	Fields show skin depth attenuation in a conducting medium, radiation and inverse linear law.	(2-4), (2-20)

Table 2-3 : Field classification based on distance of field point.

⁷ pers. comm. from B. Pease. And see, e.g., (Bannister, 1967).

⁸ ...which we will discuss in §2.3.

Additionally, we should note that the term ‘far field’ may not be ideal for describing the fields in a conductor. We will use ‘far field’ to refer to the condition $T \gg 1$, whilst noting that this could be misleading because there is no observable radiation field and that the far field components mainly comprise eddy current effects.

ii) Conductivity of Ground

Name of Approximation	Conditions	Circumstances when valid
Good conductor	$\omega \ll \omega_c$, or $\epsilon\omega/\sigma \ll 1$	Conduction current predominant
Poor conductor	$\omega \ll \omega_c$, or $\epsilon\omega/\sigma \ll 1$	Displacement current predominant
With...	$\omega = \omega_c$, or $\epsilon\omega/\sigma = 1$	
Characteristic frequency of...		

Table 2-4 : Field classification based on conductivity of ground.

iii) Other approximations

Since the transition distance is only a nominal term, created to aid the simplification of the expressions, it is instructive to ask at what distance the far-field actually takes over. In addition to $k_0 r \gg 1$, we require the loop to look like a point source, that is, $r \gg a$, where a is the radius of the loop. It can be shown, e.g. (Clemmow, 1973) §5.4.4, that a third necessary condition, is

$$r \gg k_0 a^2 \quad (2-23)$$

which is known as the *Rayleigh distance*. This condition contains both the previous ones when $k_0 a \gg 1$ and is redundant if $k_0 a \ll 1$. It gives the least distance at which the radiation field ‘takes over’ when the source cannot be treated as a single point. The Rayleigh distance is more significant than the transition distance for many problems involving microwave radiation, but it would appear to be mostly irrelevant for LF work.

At distances exceeding the *Sommerfeld distance*, of

$$1 / (k_0^3 \delta^2) \quad (2-24)$$

the usual expression for the radiation field in air breaks down due to the drain of power into the ground. The Sommerfeld distance is only significant for the propagation of ELF waves in the earth’s atmosphere (Burrows, 1978).

2.2.4 Derivation of an Optimum Frequency

For a communications system that uses induction loop antennas, and which operates mainly by magnetic induction, we know (from Faraday), that the induced signal will be less at low frequencies. However, we also know, from the above discussions, that there will be increased signal attenuation at higher frequencies due to the skin effect. We can infer that there should be an optimum frequency, and therefore a *bandpass window*, where path attenuation is minimised.

Interestingly, we can also demonstrate an optimum frequency for the co-planar field without needing to invoke a magnetic induction effect. This phenomenon is referred to, in the literature, as the *low frequency window* (Gabillard, Degauque and Wait, 1971) and is analysed in appendix A2.4.

The optimum frequency will also depend on the atmospheric noise spectrum, and we must not lose sight of the fact that an over-riding concern may be to find a quiet part of the spectrum – the LF band contains much military, direction-finding and telemetry traffic. A full discussion of optimum frequency will require a more exact model that describes an infinite half-plane conductor, but we can get an appreciation of the phenomenon by using the simpler model currently being discussed.

It will be shown in chapter 3 that an induction loop receiver generates a voltage of $\omega\mu_0 H \cdot \Phi \sqrt{R}$, where R is the resistance of the loop and Φ is a figure of merit for the induction loop, which we refer to as the specific aperture. Using this, we can write the voltage induced in a coaxial receiver loop, from (2-18), as

$$U|_{\theta=0} = \Phi \sqrt{R} \frac{m_d}{2\pi\sigma} \frac{1}{r^5} T^2 Q(T) \quad (2-25)$$

with $Q(T) = 2e^{-T} \sqrt{T^2 + (1+T)^2}$

As the frequency tends to zero, the skin depth becomes larger and $T \rightarrow 0$, so the induced voltage tends to zero. Physically this is because there is no magnetic induction at zero frequency. Similarly, at high frequencies T becomes larger, but as $T \rightarrow \infty$ we will have $\exp(-T) \rightarrow 0$ and there is also no signal. Physically this is because there is a greater degree of skin-depth attenuation, and because we are moving into the far-field, where there is no co-axial field term. The optimum value of T can be found by solving $d/dT \{T^2 Q(T)\} = 0$ to give

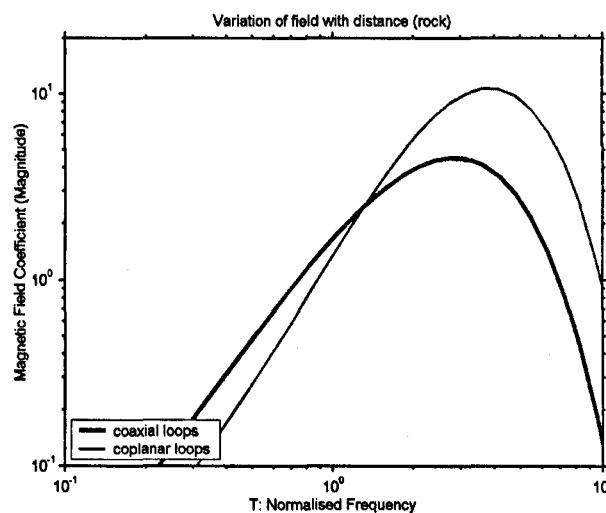
$$T^3 - 2T^2 - 2T - 1 = 0 \quad (2-26)$$

which we can solve⁹ to give $T = 2.83118\dots$, which is about 0.1% greater than $\sqrt{8}$. We could deduce from this that the optimum spacing for co-axial loop antennas is therefore about 2.8 skin depths or, from (2-15), about 0.45 of the wavelength in the conducting medium. A similar result can be obtained for co-planar loops. The data is plotted in *Figure 2-9* below. The window that Gabillard *et al.* considered to be a low-pass function is now seen to be a band-pass function. We can define a -3dB cutoff frequency, denoted by T_c and ω_c , tabulated in *Table 2-5* below, which also shows the Q -factor for the band-pass responses.

9. Solving the cubic equation algebraically results in $T = \frac{2}{3} + \frac{1}{3} \left[\sqrt[3]{\frac{79 + 3\sqrt{249}}{2}} + \sqrt[3]{\frac{79 - 3\sqrt{249}}{2}} \right] = 2.83118\dots$

As a brief example, with a co-axial induction system transmitting over 100m in a medium of conductivity 0.01S/m, we would have an available bandwidth of 21kHz centred at 20kHz. In seawater ($\sigma = 4/\text{m}$) the bandwidth would be 13Hz over 200m, centred at 13Hz.

We will return to the subject of an optimum frequency after we have considered a more accurate model of the propagation.



*Figure 2-9 : Optimum Frequency Window.
The y-axis shows the coefficient $T^2 Q(T)$, as defined in equation (2-25)
The induction effect results in a distinct band-pass response.*

co-axial loops:	-3dB b/w at $T_c = 2.88$	Peak 4.50 at $T = 2.83$	$Q = 1.0$
co-planar loops:	-3dB b/w at $T_c = 3.35$	Peak 10.7 at $T = 3.86$	$Q = 1.2$

Table 2-5 : Bandwidth of Optimum Frequency Window.

2.3 A Model for a Uniform Half-Plane

In the previous section, we considered the fields from a magnetic dipole in an infinite conducting medium, and we can assume that this is a good basic model for communications over distances that are small compared with the extent of the medium. However, if the medium is not infinite in extent we will need to impose boundary conditions for its evaluation. We could surmise that this would be required for the dual conditions that

- i) Communication is over a distance that is significant in comparison to a skin depth, and
- ii) Either a) the communications path crosses the air-rock boundary
or b) the transmitter or receiver is within a skin-depth of the air-rock interface.

In the latter case, if the transmitter and receiver were both in air, we would expect induced ground currents to influence the propagation. Likewise, if the transmitter and receiver were both buried, we could expect a significant proportion of the field to travel by an 'up over and down' path which would suffer less attenuation than a direct path through the rock.

The general solution to this problem is complicated, although it often possible to simplify the treatment. For example, we have already discussed the trivial case where the field point is sufficiently close to the source that we can assume a quasi-static approximation. However, as we have seen, sub-surface communication in caves and mines does not take place solely in the near field, but in the transition zone, when the analysis is more difficult. There are also several different antenna orientations to consider.

A convenient starting-point for an analysis – other than to assume a completely uniform earth, as we have used previously – is to assume that the medium is horizontally stratified, with the layers homogenous and isotropic. Horizontal stratification is often the case for caving applications where, in the UK, limestone may be covered by an overburden of saturated peat or underlain by a water table. It is also true for many geophysical and mining situations. A method of applying boundary conditions to a multi-layered earth was derived by (Wait, 1951; 1962; 1982). More complicated problems that have been analysed include those of tilted ground, lateral changes in conductivity, and isolated conducting ore bodies. References to these are listed in (Durkin, 1983).

A simple two-layer model would be appropriate for an analysis of a cave or mine communication system because the rock is often stratified or covered with a conductive overburden. However, the precise mode of propagation is often difficult to predict. It is known, for example, that conduction can take place along rock strata or mineralised faults; and that the presence of ore-bearing rock can distort the magnetic field lines used for radio-location. Anisotropic effects have also been observed above cave-bearing rock; these were described by (Drummond, 1987a; c) and reprinted in (Drummond, 2003). Drummond's observations are geophysically significant and would bear further investigation.

We conclude that a simple model for a uniform half-plane earth will be sufficient to allow us to make general observations about the propagation and we surmise that a more complicated model, unless developed for a specific situation, may serve only to confuse the issues at this stage of the analysis. We will show below that, in some respects, the previous 'infinite full space' model is also a good approximation.

2.3.1.1 *Boundary Conditions and Computer Simulation*

For the case of a uniform earth, the boundary conditions are straightforward to state, but a solution that requires only sine and cosine functions in 2D Cartesian co-ordinates will require

Bessel or Legendre functions in cylindrical or spherical co-ordinates; and the symmetry of a small antenna may dictate such a polar co-ordinate system. Introducing a conducting medium requires the integration to infinity of the product of a complex variable and a Bessel function. Fortunately, for the configurations we will be considering, the integrand quickly tends to zero, making the numerical integration less prone to errors.

The fields from a buried horizontal transmitter (VMD) were derived in (Wait, 1969; 1971). At that time, computer iteration would have been slow and expensive compared with the simulations we can now run on a personal computer. An MS-DOS batch program, producing tabulated data, was used by Shope to evaluate Wait's field expressions (Shope, 1991) and, more recently (Lippold, 2000b) describes how he solved Wait's equations using the *MathCad* package on a PC. For this thesis, the equations have been modelled using *MatLab*¹⁰. With a powerful 'desktop' tool such as this, we can investigate field behaviour at a level of detail that was previously not possible.

2.3.2 Orientation of Transmitter and Receiver

Because of the effect of the secondary induced field, and a reflection from the conducting ground surface, the orientation of the antenna affects the field pattern. This is evident from (Burrows, 1978) §2.4, where the surface fields for four elemental antennas – horizontal and vertical magnetic dipoles (HMD, VMD) and horizontal and vertical electric dipoles (HED, VED) – are considered. We can note, for example, that the combination of a HED and its reflection in the ground (*Figure 2-10*) results in the fields almost cancelling and there being very little radiation. For long-range VLF transmission a vertical monopole (VED) is clearly a much better radiator than a horizontal wire.

Burrows' results do not directly concern us, since he was considering the surface fields for a surface antenna but they should lead us to expect that a horizontal loop might react with a horizontal earth in a different way to a vertical loop¹¹. This leads to a number of antenna configurations that we need to consider.

10 Published by The Mathworks, www.mathworks.com

11 We can note that the Z_0 term (impedance of free space) in the field expression for an isolated HED becomes η , the *surface impedance*, when the antenna is in the vicinity of a conducting ground plane. For a 73kHz antenna (the frequency of the UK amateur VLF allocation) over conductive earth ($\sigma = 0.01\text{S/m}$, $\epsilon_r = 10$) the surface impedance is 24Ω and a HED is only 6% as efficient as a VED.

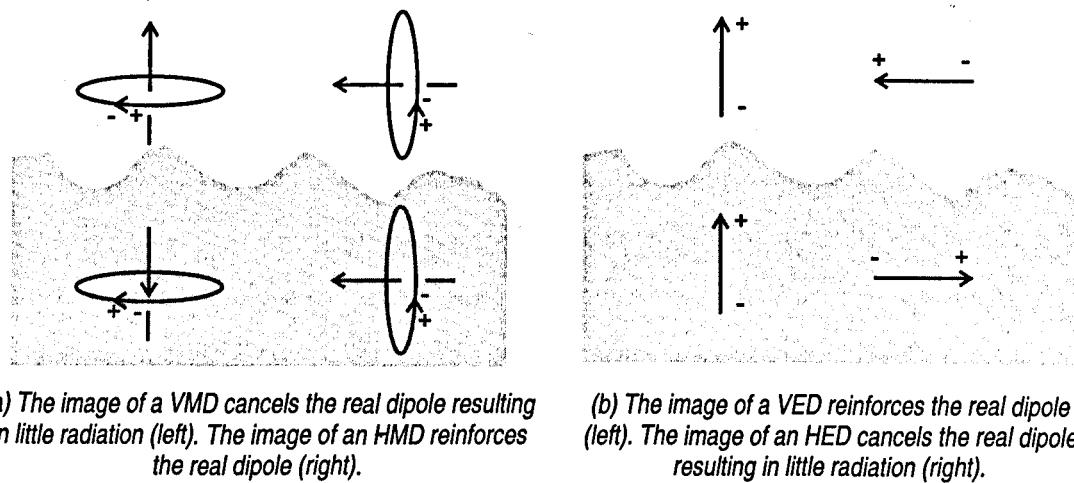


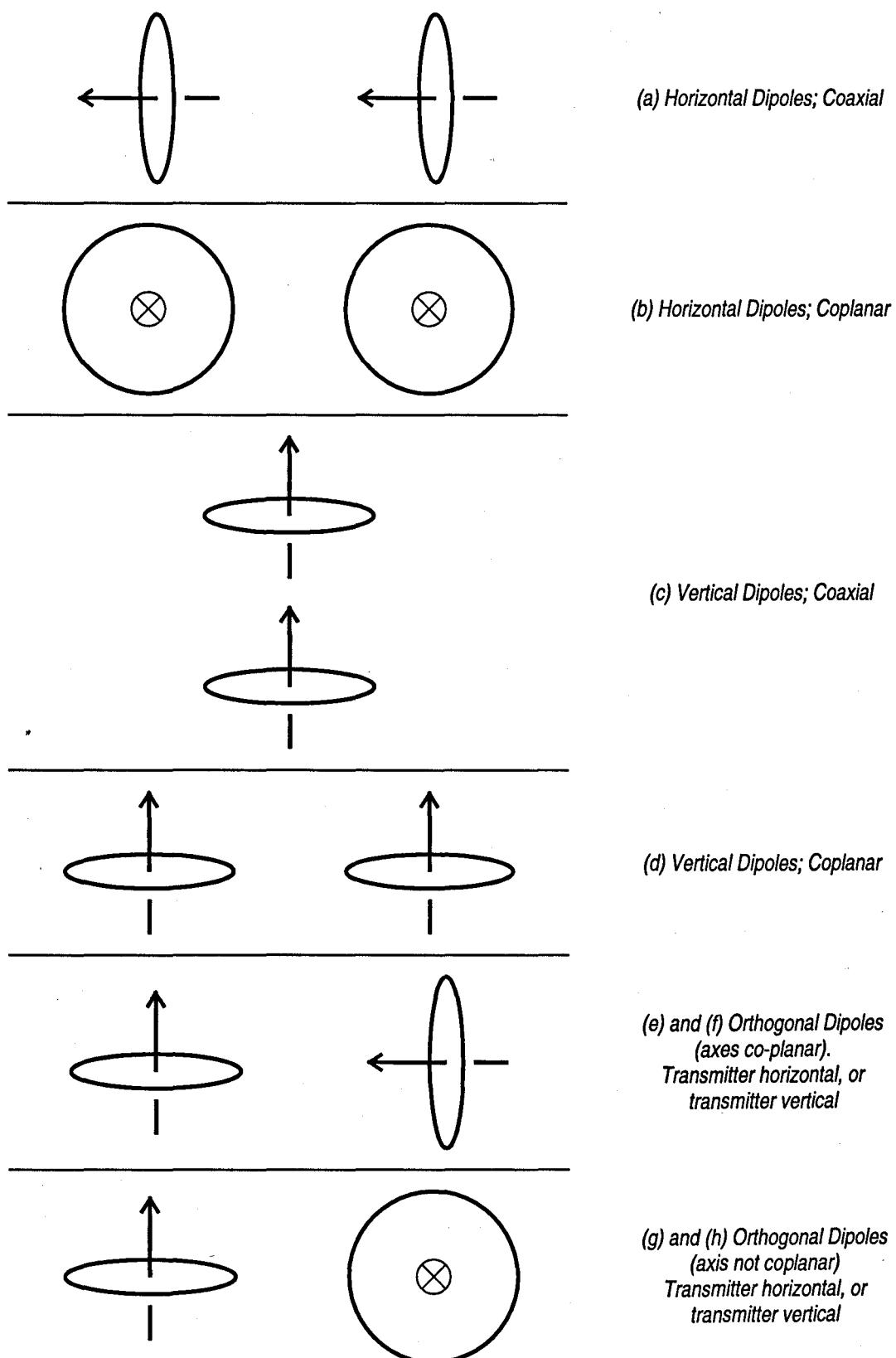
Figure 2-10 : Reflections of elemental dipoles in a conducting half-space.

2.3.2.1 Classification of Antenna Alignments

For each of the two transmitter alignments – vertical or horizontal – there are four basic alignments that the receiver can take, so there are eight possible orientations of transmitter and receiver that we must consider (*Figure 2-11*, below) if we are to present a comprehensive review. Clearly the coplanar and coaxial alignments are simply special cases of the general transmitter-receiver alignment we would normally adopt; but there are also two orthogonal arrangements.

In an infinite medium (whether conducting or not) symmetry dictates that the magnetic field must be perpendicular to a receiver loop that is coplanar to the transmitter. That is, a receiver with its axis in the plane of the transmitter loop and directed towards it (*Figure 2-11e*, below) will detect a null. However, experimental work tells us that this is not true if the loops are close to the surface of the ground. In this situation, the horizontal transmitter (VMD) will induce a signal in an orthogonal receiver loop, due to the *secondary field* and this can be the most significant of the field components. Even if our modelling task were simply to obtain an estimate of the range of transmission we can deduce that a simple ‘full space’ model (that did not demonstrate this behaviour) would be insufficient. Additionally, for geophysical investigation (e.g. measuring ground conductivity by a non-contact means) and radiolocation, where the exact shape of the field lines is important, a more complex model is clearly needed.

The second orthogonal case is where the transmitter and receiver axes are orthogonal and not coplanar (*Figure 2-11g,h*, below). Symmetry dictates that this configuration should always detect a null, but this assumes that the ground is isotropic.

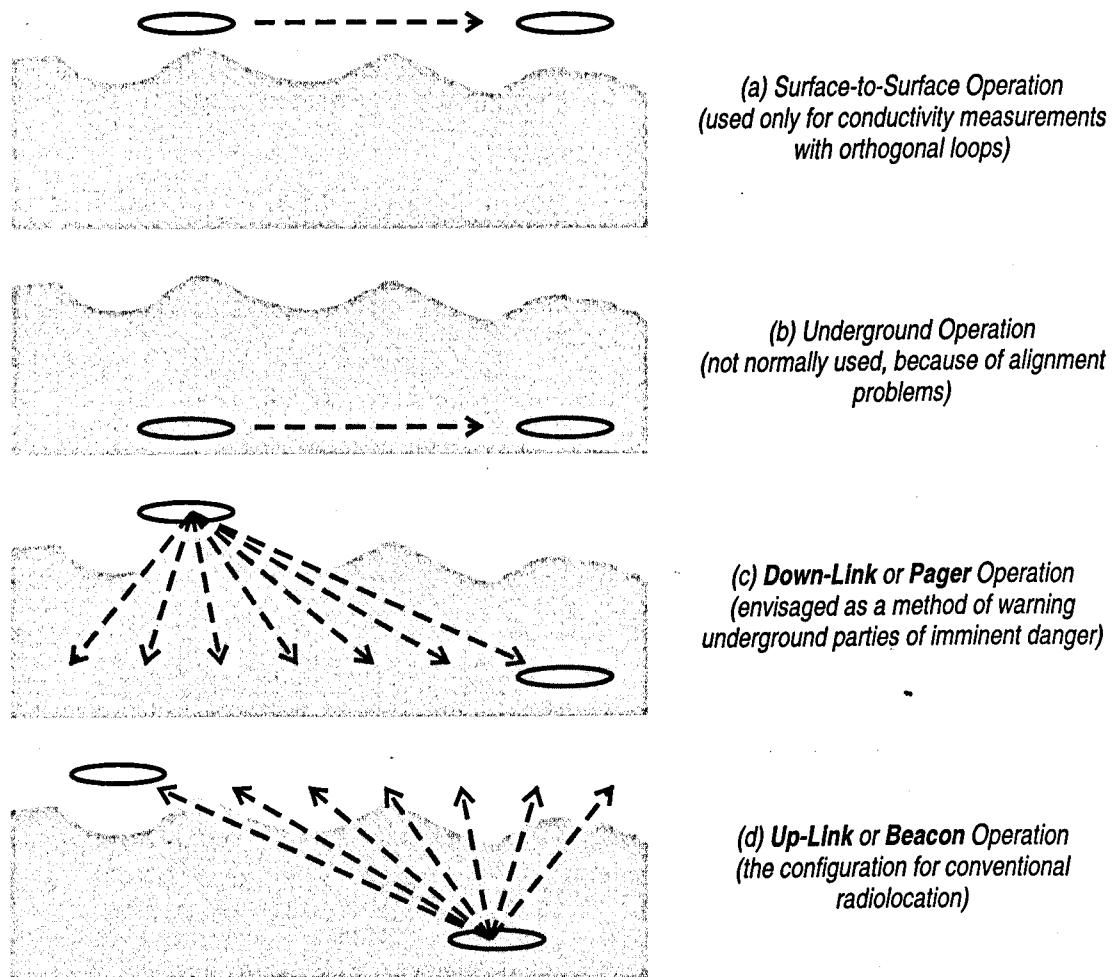


*Figure 2-11 : Eight basic alignments of transmitting and receiving induction loops.
The field characteristics over a horizontal earth depend on whether the transmitter is vertical or horizontal.
For each of these two transmitter alignments there are four basic alignments that the receiver can take.*

2.3.2.2 Classification of Communications Paths

As well as the basic orientations of transmitter and receiver shown in *Figure 2-11* we can identify four types of communication system, based on whether and how the path crosses the air/rock interface (*Figure 2-12*, below). We could expect – and it will be shown later – that the fields are not reciprocal. That is, the up-link field is different to the down-link field. In addition, of course, in a practical system, we would expect down-link transmission to be easier because of the attenuation, with distance, of atmospheric noise.

A complete analysis of all possible antenna systems is beyond the scope of this thesis, and may form the basis for future work. Additionally, we can note that several ‘exotic’ antenna systems are possible, including one where the magnetic dipole is replaced by a so-called *anapole* or toroidal antenna – this is discussed briefly in §4.8.



*Figure 2-12 : Classification of communications systems, based on signal path.
The diagrams show horizontal loops but each system can utilise any of the configurations of
Figure 2-11.*

2.3.3 Field Expressions

2.3.3.1 Horizontal Loops (VMD) in coaxial alignment

It is common for induction systems in both cave and mine applications to use horizontal loop antennas, arranged co-axially as in *Figure 2-11c* with one antenna underground and the other on the surface of the ground. This alignment is easy to achieve¹² and results in the shortest distance between the antennas, and therefore the maximum signal. It is also a good alignment to adopt for pager and beacon operations (*Figure 2-12c, d*), and for radiolocation.

For the conditions where the ground is a semi-infinite, homogenous medium and is a good conductor, J.R. Wait derived expressions (Wait, 1969) for the field due to a buried infinitesimal loop, which were summarised in (Wait, 1971). In terms of a dummy variable of integration, λ , the expressions were given as

$$(H_p, H_z) = \frac{m_d}{2\pi} \int_0^\infty d\lambda \frac{\lambda^3}{\lambda + u} e^{-\lambda z} e^{-uh} (J_1(\lambda\rho), J_0(\lambda\rho)), \quad \text{with } u^2 = \lambda^2 + j\omega\mu\sigma \quad (2-27)$$

where the transmitter loop is buried at a depth h , the receiver is a distance z above the surface, at an offset of ρ , and H_p, H_z are the radial and vertical components of the magnetic field. Wait demonstrated that a change of variable to

$$x = \lambda h \quad (2-28)$$

allowed the integrals to be expressed in a more convenient, dimensionless, form. For this, we define

$$T = \frac{h}{\delta}, \quad D = \frac{\rho}{h}, \quad Z = \frac{z}{h}, \quad \text{with } \delta = \sqrt{\frac{2}{\omega\mu\sigma}}. \quad \text{Also } F = \sqrt{2} \cdot T \quad (2-29)$$

where the need for the F term arises from Wait's and Shope's normalisation of depth, as discussed in §2.2.3.1, p37. We retained this definition to allow a direct comparison between our graphs and those of Shope, although it does give rise to scope for confusion.

We now define two dimensionless complex functions, P and Q such that the magnetic field can be written as

$$\mathbf{H} = \frac{m_d}{2\pi h^3} \{ P \hat{\mathbf{p}} + Q \hat{\mathbf{z}} \}. \quad (2-30)$$

The quantities P and Q can now be expressed in terms of the dummy variable x as

$$(P, Q) = \int_0^\infty dx \frac{x^3 e^{-xz}}{x + U} e^{-U} (J_1(xD), J_0(xD)), \quad \text{with } U^2 = x^2 + jF^2 \quad (2-31)$$

12 See §2.4.5 for a further discussion of radiolocation.

which can conveniently be evaluated using *MatLab* or similar software for mathematical analysis.

2.3.3.2 Modification for Down-Link Fields

Wait's formulations were for an up-link field, because he was considering location devices for trapped miners. In (Durkin, 1983), during a discussion of communications systems for miners, the corresponding down-link expression is derived, namely

$$(P, Q) = \int_0^{\infty} dx \frac{x^3 e^{-xz}}{x+U} e^{-U} \left(J_1 \frac{U}{x} (xD), J_0 (xD) \right) \quad (2-32)$$

showing that the horizontal field (P) does not obey a 'reciprocity' rule. Both Durkin and Wait have considered the situation where the transmitter loop is of finite size, when it can be modelled by an additional Bessel function term. We have omitted this term, and will assume that the loops are infinitesimal, or that the field point is sufficiently far for the loop to behave as a point source.

2.3.3.3 Vertical Loops (HMD) in Co-planar Alignment

Although horizontal loops are easy to use, vertical loops are only marginally less so – especially if two orthogonal loops are driven in phase quadrature to achieve a circularly polarised transmission. In addition, there is the advantage of a higher signal strength in a coplanar arrangement, as shown in *Figure 2-7* and *Figure 2-9*. The field due to a buried HMD is derived by (Wait, 1972).

We suggested some time ago that HMDs may couple to the conducting medium in a different manner, and (Drummond, 1987b) has also pointed out the benefits of a coplanar alignment, but this alignment has not been adopted for cave or mine communications. Whilst an analysis of the field propagation from a horizontal dipole is undoubtedly important, it has not been studied for this thesis, and remains a subject for future work.

We shall note, in chapter 6, that the atmospheric noise is likely to propagate vertically downwards through the ground, and so a HMD is likely to detect more noise than a VMD. This would also be an influence on the choice of antenna orientation.

2.3.3.4 Elevated Source and Field Point

Wait's results, referred to above, are for the fields due to a buried source. In (Wait, 1982) §3, expressions are given for the fields due to an elevated source above a multi-layered earth. The results are tabulated in a form that separates the primary and secondary fields, and includes results for horizontal dipoles. A useful future study would be to extend this tabulation to cover all the antenna orientations described in §2.3.2.

2.3.3.5 Orthogonal Loops – Study of Secondary Field

The orthogonal arrangement of *Figure 2-11e*, can be considered as a special case of the field from a VMD transmitter. The main interest in this configuration is when both loops are at ground level when, in this orientation, the receiver will detect only the secondary field. This arrangement can be used to derive the conductivity of the ground and, although this is important in the study of sub-surface communications it will not be addressed further in this thesis. However, it is useful to consider, briefly, the situation where a shallow-buried underground transmitter is communicating with a receiver at a large offset from the ‘ground zero’ point. In this situation it can be shown that the secondary field – detected by the orthogonal antenna – can be considerably greater than the primary field – detected with a coplanar receiver. This has been observed practically (Drummond, *pers. comm.* to B. Pease) and lends weight to the suggestion that the orientation of the antennas in a communications system is important.

2.3.4 Computer Simulation of a Buried VMD

We have presented expressions – (2-30) and (2-31) – for the horizontal and vertical fields from a buried horizontal loop, or vertical magnetic dipole (VMD). These can be solved with the aid of a computer program and the results presented graphically. We have chosen to use the program MatLab for this task. There are several reasons for investigating the fields at the level of detail now presented; namely,

- i) We hope, by inspection, to be able to deduce whether the simple full-space model described in §2.2.3 is an adequate approximation to the more exact model presented in §2.3.2.2. If this is so, then we will have a simpler method of predicting communications range and optimum frequency.
- ii) We can investigate the phase of the secondary field in some detail, and this will help us in a study of radiolocation, which can be prone to errors at skin-depth distances.
- iii) A study of the fields may reveal unusual behaviour, which we can further investigate with the channel sounder. We already know that the up-link path is not identical to the down-link path; and there may be other observations we can make on, for example, ideal antenna orientation.

A summary of the salient points of the MatLab computer programs that were used to investigate the field characteristics are given in appendix A2.5. The programs are indexed in appendix A9.

2.3.4.1 Comparison of half-space and full-space models

Figure 2-13, below, shows the ratio of the field derived by the simple whole-space model to that of the more accurate half-space model, for the case of the co-axial arrangement, described in §2.3.3.1. In essence, it shows the ratio of the vertical field given by (2-25) to that given by (2-30). We can see that at three skin depths the more accurate half-space model gives a field strength that is 85% of that predicted by the simpler full-space model. We must bear in mind that this result only applies to the coaxial field direction.

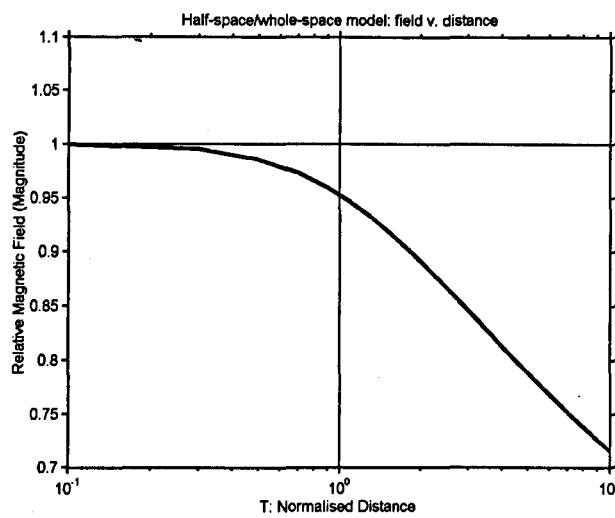


Figure 2-13 : Comparison of full-plane and half-plane models.
The field strength for this half-plane model is that of a VMD on the surface of the ground, and a co-axial sub-surface field point (or vice versa).

2.3.4.2 Q-field v. offset D for a Range of Ground Conductivity, \mathcal{T}

Figure 2-14 shows the vertical component of the magnetic field at the surface of the ground due to an underground transmitter (horizontal loop). The graphs show the field versus the horizontal offset from ground zero (GZ) for a range of ground conductivity. The offset, D , is normalised with respect to depth, h and the ground conductivity, \mathcal{T} , is expressed in skin depths, see equation (2-29). We observe that for a non-conducting medium ($\mathcal{T} = 0$) there is a null in the vertical component of the field at an offset of $D = \sqrt{2}$. This is apparent from the near-field equation, (2-5), by setting $\tan\theta = \sqrt{2}$. In a non-conducting medium the phase of the field changes sharply by 180° at $D = \sqrt{2}$ which is also as we would expect – the field line curves over, becomes horizontal at $D = \sqrt{2}$ and changes direction, curving down; the change of phase representing the change in direction.

For $\mathcal{T} > 0$, the null in the vertical field disappears, and the phase change at $D = \sqrt{2}$ smooths out, which is evidence of a secondary field with a different phase and direction to the primary field. *Figure 2-15* shows that the null starts to be filled in at quite low values of \mathcal{T} .

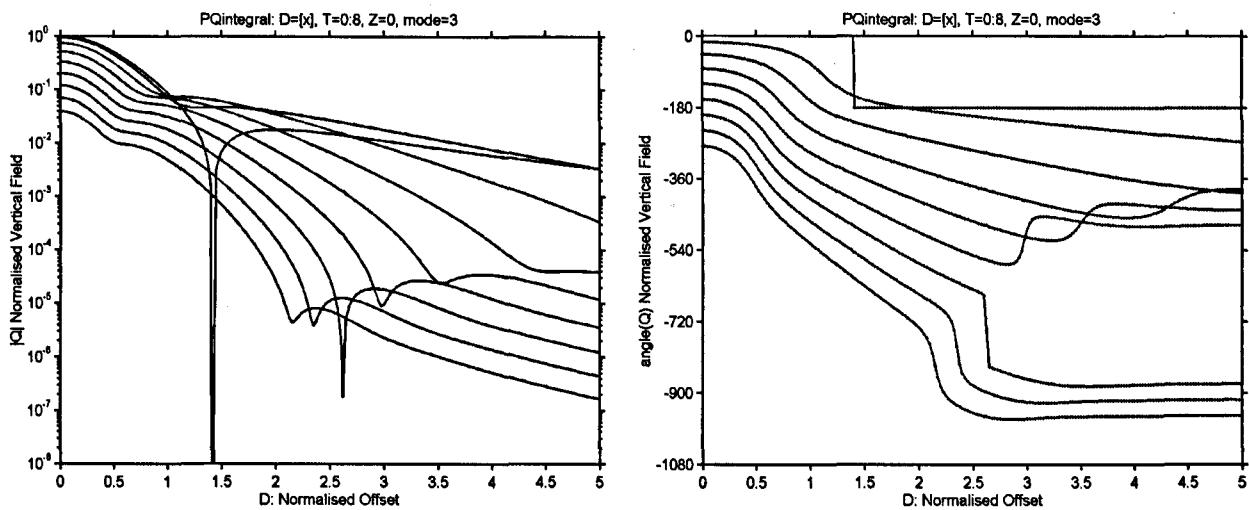


Figure 2-14 : Q -field v. normalised offset D .
Magnitude (left) and phase (right), for $\bar{T}=0$ to 8 reading from top to bottom of graphs.

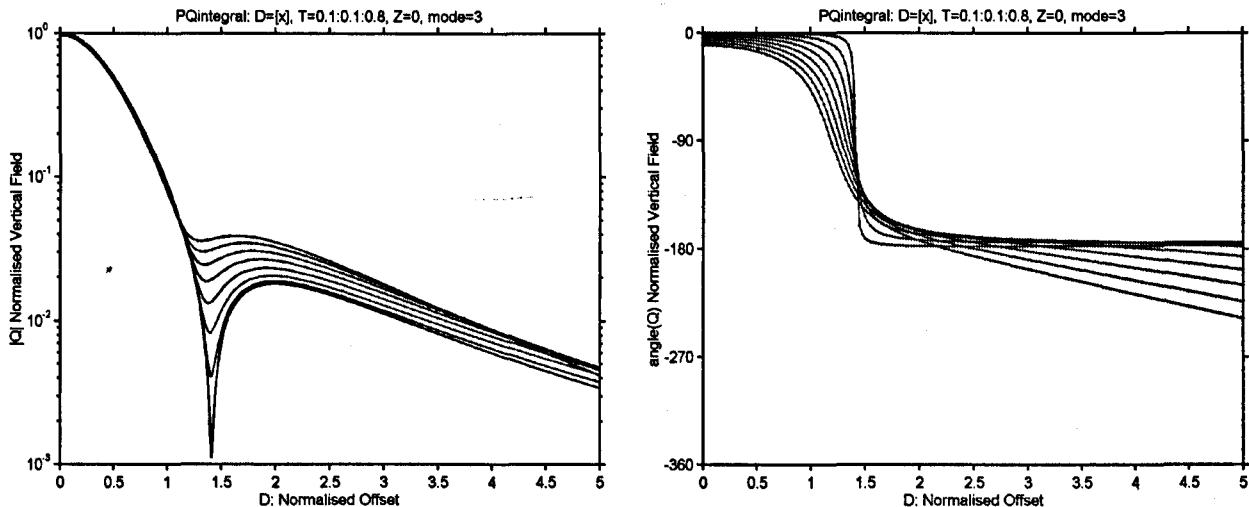


Figure 2-15 : Q -field v. normalised offset D (detail).
Magnitude (left) and phase (right), for $\bar{T}=0.1$ to 0.8 in steps of 0.1.

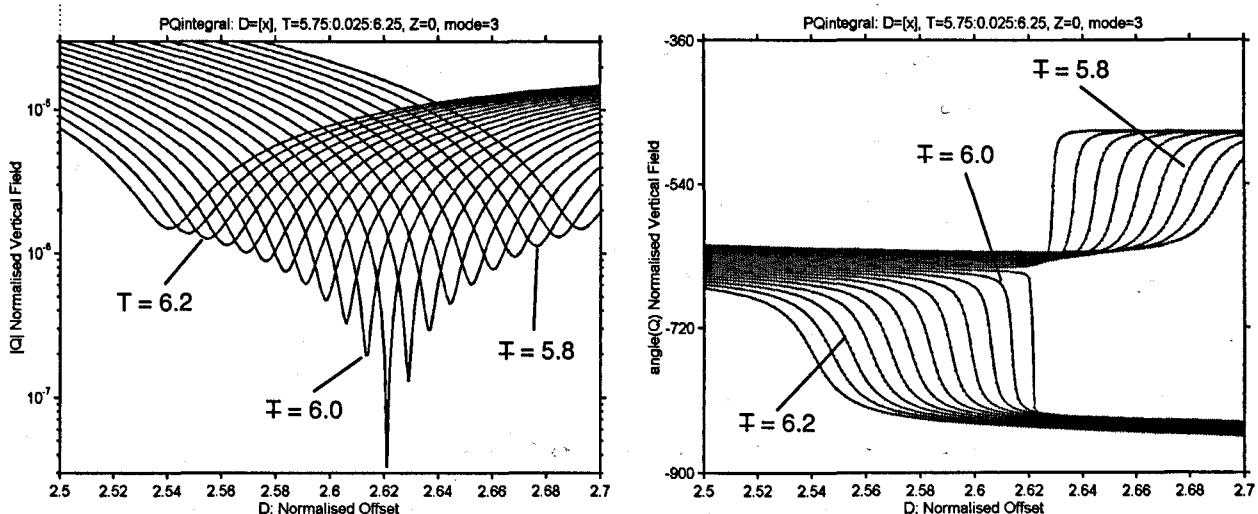


Figure 2-16 : Q -field v. normalised offset D (region near $\bar{T}=6$).
Magnitude (left) and phase (right), for $\bar{T}=5.75$ to 6.25 in steps of 0.025.

The Matlab function we have used to calculate the phase returns an angle of between -180° and $+180^\circ$. On the assumption that the phase must always be negative, and must vary smoothly, the program includes a routine to handle the abrupt changes of phase that the function returns. It also causes the phase at $D = 0$ to decrease smoothly with increasing \mathcal{F} . However, the phase does display some unexpected behaviour, notably at around the $\mathcal{F} = 6$ value, where there appears to be a null in the magnitude and an abrupt change in phase of 180° . It is not understood whether this is an artefact of the computer iteration, or a real physical effect. The fact that the null and the phase change occur together suggests that this is a real effect. Moreover, the effect has been observed by the author in programs on different platforms, written by (Shope, 1991) and (Lippold, 2000b). It is conceivable that the combination of the primary and secondary field could give rise to a horizontal field line at this location, but further investigation will be needed. However, this phenomenon may be more of academic than practical interest because it is only observable at a specific value of \mathcal{F} (i.e. frequency), and under conditions where the overall field strength is extremely low. Additionally, of course, it depends on the ground having a uniform conductivity, and this is unlikely to be the case.

The detail of this region is shown in **Figure 2-16** where the deepest null can be seen to be at $\mathcal{F} = 5.975$. The phase can be seen to be approaching an abrupt change of 180° at this value of \mathcal{F} .

2.3.4.3 P-field v. offset D for a Range of Ground Conductivity, \mathcal{F}

Figure 2-17 shows the horizontal component of the magnetic field at the surface of the ground due to an underground transmitter (horizontal loop). As in the previous section, the graphs show the field versus the horizontal offset from ground zero (GZ) for a range of ground conductivity. It is interesting to observe that beyond a certain offset, the phase appears to remain constant. In fact, a similar behaviour is evident for the Q field, but we can surmise that

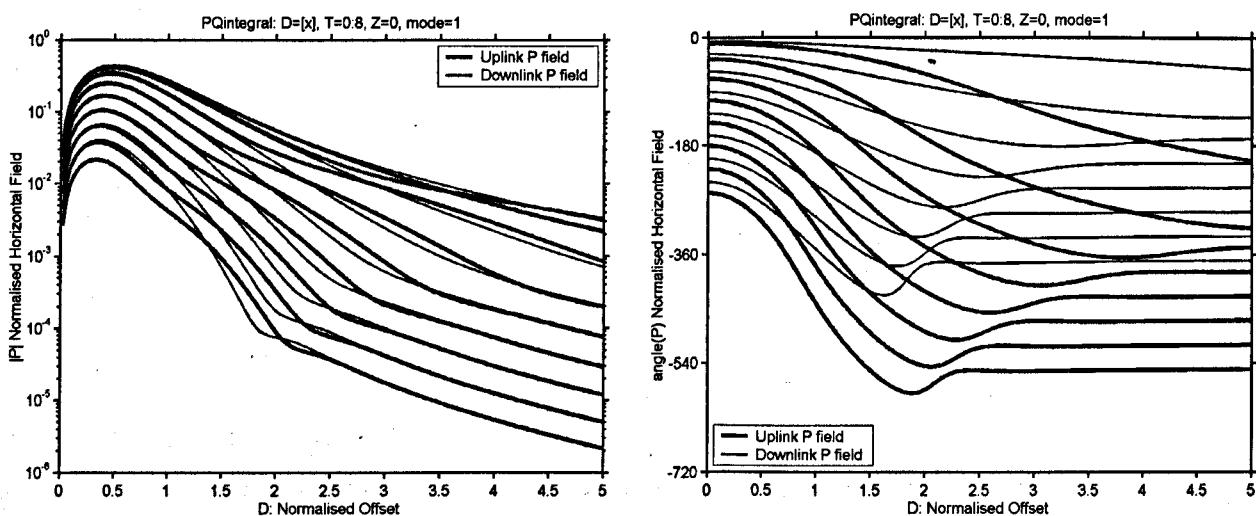


Figure 2-17 : P-field v. normalised offset D.
Magnitude (left) and phase (right), for $\mathcal{F}=0$ to 8, lines reading top to bottom.

giving a physical interpretation to this behaviour – and confirming that it is a feature of the model and not an artefact of the iteration – could be difficult.

Of more interest than the P field alone, is the ratio of P to Q since this is used to measure the field angle, which is used in one form of radiolocation. This will be mentioned, briefly, in the next chapter, and will be the subject of future study.

2.3.5 Optimum Frequency using the Half-Plane Model

Using the half-plane model we can return to the question of choosing an optimum frequency. The data from *Figure 2-9*, obtained using the previous, simpler, model is re-plotted with the x -axis containing frequency, rather than the normalised parameter T . This gives us a set of graphs for different depths, and the locus of the maxima follows a line corresponding to the optimum value of T . This data is shown in *Figure 2-18* for two values of ground conductivity.

The corresponding data for the more accurate half-plane model, using Wait's Sommerfeld integral, is superimposed and shows that there is a close agreement. For a given depth, frequency and conductivity, the approximate model results in a slightly greater signal attenuation. *Figure 2-19* shows that if the value of T used in the simple model is reduced by 5% there is a very close match to the full model, and so we can say that the optimum frequency, instead of being at $T \approx 2.83$ is at $T \approx 2.68$.

We saw from *Figure 2-9* that the optimum frequency for a coplanar alignment was different to that for the coaxial alignment. We have investigated this; as well as the situation where the field point is elevated above the ground by a short distance. These results were presented in (Gibson and Darnell, 2000) in which it was shown that there was some variation in optimum frequency under these conditions. These results will be further assessed in future work.

Noting this variation, and the low Q -factor exhibited in the graphs, we can conclude that the optimum frequency need not be rigidly defined; but we should aim to operate at a distance of around 2.8δ or 0.45λ if we are using a narrow-band system. This does not mean, of course, that the signal gets greater as the distance from the transmitter increases; but that, at a given distance, the signal is greatest if the frequency is such that the field point is at 2.8δ or 0.45λ . The signal/noise ratio may well be optimal at a different frequency, depending on the background spectrum of the noise.

This evaluation of optimum frequency can easily be extended to cover other orientations of transmitter and receiver. A simpler evaluation of Wait's expression was undertaken by Caffey (1978) who also showed that the maximum induced signal between co-axial coils occurred at a burial depth of "almost half a wavelength". The optimum frequency for communications

contrasts with the requirement for good radiolocation, which is $T \ll 1$ so that the field lines are not distorted by far-field effects, as mentioned briefly in §2.4.5.

Given a 'profile' for the variation in atmospheric noise with frequency, we can adapt *Figure 2-18* and *Figure 2-19* to show the signal to noise ratio rather than the received signal voltage. This will be discussed further in chapter 6.

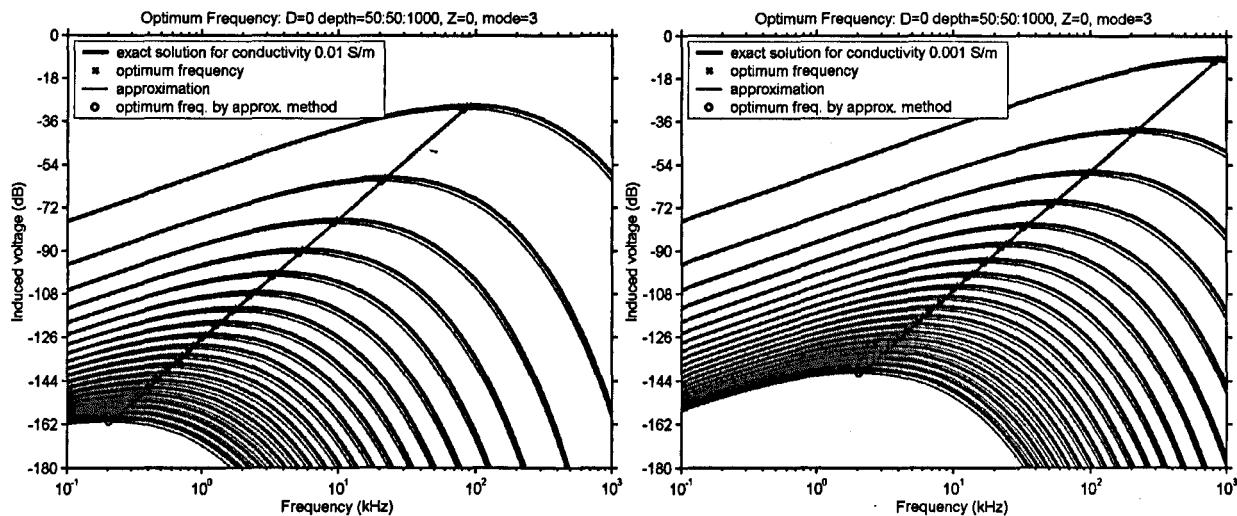


Figure 2-18 : Induced voltage v. frequency, co-axial antennas.

Antenna orientation: surface station and sub-surface station.

Graphs show the exact solution (due to Wait) and a simpler uniform space model. The straight lines are the locus of the optimum frequency for different depths.

Left: ground conductivity 0.01S/m. **Right:** ground conductivity 0.001S/m.

Depth: (top to bottom) 50m to 1000m in steps of 50m.

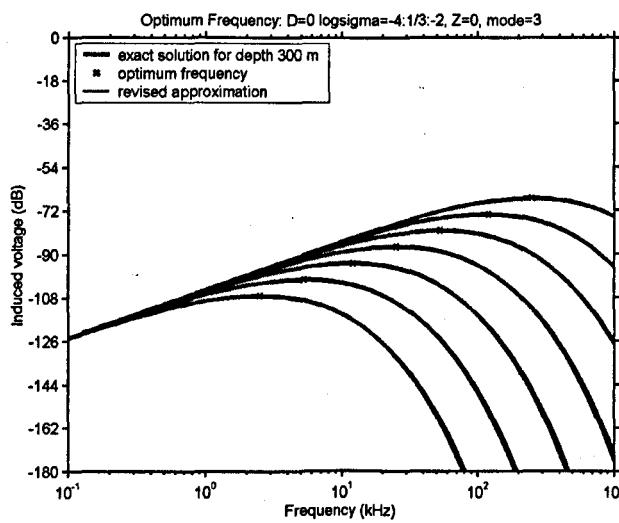


Figure 2-19 : Induced voltage v. frequency, comparison of models.

The graphs show the exact solution (due to Wait) and the simpler uniform space model, with T set to 95% of its true value.

Depth: 300m. Conductivity: (left to right) [0.01, 0.005, 0.002, 0.001, 0.0005, 0.0002, 0.0001] S/m.

2.4 Other Aspects of Propagation

2.4.1 Choice of Antenna System

In this chapter we have considered the fields from a small magnetic dipole, which is easy to model; however, this is just one of several types of antennas that can be used. Antenna systems will be discussed in chapter 5. We acknowledge that in a fully comprehensive study of sub-surface communications we would have to consider a range of antenna systems, some of which are difficult to model – for example, the current-injection antenna described by (Lippold, 2000a). Although this antenna is a key feature of modern cave rescue systems, we can attempt to justify our omission of its detailed discussion by restricting our analysis to portable communications systems using small antennas – the concept of ‘pagers’ and ‘beacons’.

2.4.2 Dispersion

In our analysis, we have assumed that the quantities μ , ϵ and σ do not vary with frequency. However, we have also seen (§2.2.2.1) that the speed of propagation is proportional to \sqrt{f} in a conducting medium. This effect will give rise to the dispersion of a broadband pulse. We will have to bear in mind this effect when we consider the synchronisation of our channel sounder, to be discussed in chapter 8.

The use of dispersion as an aid to cave surveying was discussed in any undergraduate report by (Walters, 1983), in which he concluded that it was unsuitable, largely because the degree of dispersion depends on the direction of propagation. We could investigate this further by a close analysis of the graphs in *Figure 2-3* and *Figure 2-5*.

2.4.3 Multipath Propagation

The ratio of the wavelength in rock to the wavelength in air is

$$\frac{2\pi\delta}{c/f} = \sqrt{\frac{2\epsilon\omega}{\sigma}} = \sqrt{\frac{2\omega}{\omega_c}} \quad (2-33)$$

where ω_c is the characteristic frequency, discussed earlier. The reduction in wavelength is related to how good the conductor is. Since ‘good’ means $\sigma/\epsilon\omega \gg 1$, say at least 10, then the wavelength is at most 45% of the free-space value. If $\sigma/\epsilon\omega = 1000$ then the wavelength is only 5% of the free-space value.

Because we are operating close to the transmitter there would seem to be little opportunity for multi-path effects to arise; although, if there were a possibility of such effects then the shorter wavelength and slower speed of propagation might exacerbate the effect.

Far-field noise and interference is usually considered to propagate vertically downwards through the rock, because of its high refractive index relative to air. But there may be unforeseen modes of propagation – cavers have reported being able to receive BBC Radio 4, on 198kHz, deep underground in caves in Spain, whilst the signal was undetectable at the surface.

2.4.4 Geophysical Methods

We have commented on the use of VLF induction systems as an aid to conductivity measurement. It is known that the polarisation of broadcast VLF signals changes in the vicinity of fault zones, due to the mineralisation of the fault setting up conduction paths. The same mineralisation can cause VHF radio to be conveyed some distance underground. Geophysicists make use of the polarisation change to detect fault zones (Phillips and Richards, 1975). Cavers have suggested using the same techniques, but as an aid to detecting cave entrances that might occur along fault planes. There are many examples of caves that have developed along fault lines; one local classic example being Tatham Wife Hole in the Yorkshire Dales, located on the western flank of Ingleborough.

2.4.5 Radiolocation

Radiolocation, in the speleological context, is the procedure of determining the position and depth of an underground radio transmitter beacon by making measurements on the surface using a radio receiver. The name is, perhaps, a misnomer because the underlying physical principle owes more to the phenomenon of magnetic induction than to ‘true’ radio, and the procedures are different to those used in above-ground radiolocation¹³.

In practice, we may find that an accurate field analysis is not be required for the theoretical evaluation of a communication path. However, magnetic fields are also used for underground radiolocation, for cave surveying and rescue purposes, for which knowledge of the shape of the field lines is essential. We have observed that the quasi-static model soon breaks down, and this means that conventional radiolocation is successful only at distances that are small in relation to a skin depth, which is in contrast to the optimum frequency requirement of three skin depths.

Speleological radiolocation dates, in its present form, from the mid-1950s when transistorised equipment started to become readily available to experimenters and the mid-1960s also saw a spate of research by the US Bureau of Mines. The method of radiolocation has changed little since its inception. The method is described by (Glover, 1976) and (France,

13 For above-ground use, a direction-finding antenna usually consists of a horizontal magnetic dipole that is used to detect the horizontal circular magnetic flux lines from a vertical electric dipole transmitter; and utilising the *magnetic* field from an *electric* dipole simplifies the problem.

2001) and usually involves making measurements of distance and field angle (DFA). A cruder method of depth determination utilises measurements of absolute signal strength (ASS). However, because of the need for calibration, this method is not usually used.

2.4.5.1 Radiolocation at Skin-depth Distances

We have proposed a ratiometric method, based on a measurement of the vertical component of the field gradient (VFG) that does not require calibration (Gibson, 1996b) which was investigated by USA caver Brian Pease. Pease noticed that, in the transition zone, where the quasi-static field model starts to break down, this method tended to underestimate the depth whereas the ASS method (based on the transmitter power and the inverse cube law) tended to overestimate it. Pease further observed that the mean of the two errors was low (Pease, 1997). For example at two skin depths, the error using the ASS method is +22%, whereas it is -19% for the VFG method. The mean of these errors is 2%.

We have developed this concept further and devised a technique that, in theory, results in a zero depth error. If the ground conductivity were known then a straightforward application of the full field equations would suffice, but our technique works *without* requiring the conductivity to be stated. The reason this is possible is that, in making the various field strength measurements we have *inter alia* obtained sufficient information to determine the conductivity as well as the depth.

We have extended the concept to propose a method of radiolocation that is similar in concept to global positioning systems. We feel that this is a significant area of original work, and we intend to undertake field tests in the near future. However, it lies outside the main scope of this thesis and so the reader is referred to previously published reports (Gibson, 1998; 2002). We have also given presentations on this topic at the URSI symposiums in 1998, 1999 and 2000, (see appendix A1) but conference proceedings were not published.

2.4.5.2 The 'Tilted Body' Problem

Cave radiolocation relies on the transmitter antenna being horizontal – i.e. a vertical magnetic dipole. Tilting it through a small angle does not result in a large error because of an effect that has come to be known as the 'thirds rule' (Gibson, 1998). However, the methods used by cavers will not work if the antenna is grossly tilted. This is a problem that has been addressed by the mining industry, in the context of body-worn beacons. In the event of an accident, it would be possible to locate a prone miner by means of a small body-worn transmitter, but this requires a deeper study of the problem.

This generalised location problem has been addressed by (Nessler, 1981; 1989a; b; Nessler and Bohler, 1996; Nessler and Seeburger, 1986) but only in the context of a quasi-static field.

In (Nessler, 1981) a frequency of 1kHz and a conductivity of 10mS/m are used in an example. The skin depth for these conditions is 160m, yet the author claims that the quasi-static model is valid at "under 200m", which we have shown not to be the case. In (Nessler and Bohler, 1996) a system is described at 3kHz and 1mS/m, for which the skin depth is 290m. The operating distance is 100m, for which the quasi-static model would be reasonably good, but the authors claim that the reason the model works is the distance is $1/1000^{\text{th}}$ of a wavelength and is therefore in the "extreme near field". This, of course, is incorrect because the wavelength in the conducting medium would be reduced to 2π times the skin depth. These are only minor criticisms, and the general technique that Nessler *et al.* describe is highly interesting, and we hope to investigate its application to caving problems.

2.5 Concluding Remarks

We have explained the propagation of the magnetic field from a magnetic dipole in some detail, and we have demonstrated that the field can be evaluated with the aid of a desktop computer program. We have given results only for the case of a buried horizontal transmitter, but a further study of different orientations will allow us to make predictions about the communications path that we can attempt to verify by means of channel sounding.

We have given a very brief summary of radiolocation in the context of cave and mine usage and have indicated that the computer technique used to investigate the fields from a small dipole can be applied to a study of radiolocation. Such a study is outside the main scope of this thesis although the two topics of radiolocation and communications are linked. We will not comment further on radiolocation in this thesis, but the channel sounding techniques that we will describe here could be used to further pursue this topic.

2.6 References¹⁴

- Bannister, P. R. (1967). Quasi-Static Fields of Dipole Antennas Located above the Earth's Surface. *Radio Science* 2, 1093-1103.
- Burrows, M. L. (1978). *ELF Communications Antennas. [IEE Electromagnetic Waves Series: 5]*. Stevenage: Peter Peregrinus.
- Clemmow, P. C. (1973). *An Introduction to Electromagnetic Theory*. Cambridge: Cambridge University Press.
- Drummond, I. (1987a). A Cave Radio in the Field - Summer 1987. *Speleonics* 9, 7-8.
- (1987b). Magnetic Moments #6: The Transition Zone. *Speleonics* 8, 13.
- (1987c). Surface Locations above Lechuguilla Cave, Carlsbad Caverns National Park, New Mexico, Using Cave Radio Equipment. Alberta, Unpublished Report for Carlsbad Caverns National Park (Published by CREG as Drummond, 2003).
- (2003). Errors in Ground-Zero Radiolocation at Lechuguilla. *CREGJ* 50, 24-27.
- Durkin, J. (1983). Study of through-the-Earth Communications (Phd Thesis). *School of Engineering*. Pittsburgh: University of Pittsburgh.
- Ferrari, R. L. (1975). *An Introduction to Electromagnetic Fields*. Wokingham: Van Nostrand Reinhold Co. Ltd.
- France, S. (2001). Cave Surveying by Radio-Location - 2. *CREGJ* 44, 21-23.
- Gabillard, P. D., P. Degauque and J. R. Wait (1971). Subsurface Electromagnetic Telecommunication - a Review. *IEEE Transactions on Communication Technology COM-19*(6), 1217-1228.
- Gibson, D. (1996a). Cave Radio Notebook 13: Inverse Cube Law. *CREGJ* 24, 19-20.
- (1996b). How Accurate is Radio-Location? *Cave & Karst Science* 23(2), 77-80.
- (1998). Radiolocation Errors Arising from a Tilted Loop. *Compass Points* 21, 18-20.
- (2002). Cave Surveying by Radiolocation - 3. *CREGJ* 48, 11-16.
- Gibson, D. and M. Darnell (2000). *Simulation of an Underground Communications Antenna System*. MatLab DSP & Communications Conference, Reading: The Mathworks.
- Glover, R. R. (1976). Cave Surveying by Electromagnetic Induction. In *Surveying Caves*. ed. B. Ellis. Bridgwater: British Cave Research Association.
- Lippold, R. (2000a). Computer Simulation of a Grounded Wire Antenna. *CREGJ* 42, 5-11.
- (2000b). Computer Simulation of a Small Buried Loop. *CREGJ* 39, 5-8.
- Lord, H. (1988). Letter: Radio Communications through Rock. *Electronics & Wireless World* 94(1627), p458.
- Nessler, N. H. (1981). *Subsurface Position Finding with ELF Waves*. IEE Second International Conference on Antennas and Propagation, York: IEE.

¹⁴ References to *CREGJ* are to the *Cave Radio & Electronics Group Journal*; a non-peer-reviewed low-circulation magazine published by a special interest group of the *British Cave Research Association*, bcra.org.uk. The CREG journal is filed at the British Library as ISSN 1361-4800. *Cave & Karst Science* is the BCRA's peer-reviewed cave science journal, ISSN 1356-191X. *Speleonics* is the newsletter of the Communications and Electronics Section of the USA-based *National Speleological Association* – see www.caves.org / section / commelect

- (1989a). *Location Technique for Underground Engineering Using Propagation Effects of ELF Signals*. Proc. Int. Symp. Antennas and Propagation, Tokyo.
- (1989b). Location Technique with Self-Locating Sensing Antennae. *Measurement* 7(1), 13-18.
- Nessler, N. H. and N. Bohler (1996). *Location System with ELF-Coded Signals for Underground Applications with Coded Responsive Transmitters*. Proc. Int. Symp. Antennas and Propagation, Chiba, Japan.
- Nessler, N. H. and R. Seeburger (1986). *Position Finding in Mines with ELF Waves for Arbitrarily Oriented Transmitters*. Proc. 5th Int. IMEKO Symposium on Intelligent Measurement, Jena.
- Pease, B. (1997). Determining Depth by Radiolocation: An Extreme Case. *CREGJ* 27, 22-25.
- Phillips, W. J. and W. E. Richards (1975). A Study of the Effectiveness of the VLF Method for the Location of Narrow-Mineralised Fault Zones. *Geoexploration* 13, 215-226.
- Shope, S. M. (1991). A Theoretical Model of Radio-Location. *NSS Bulletin* 53(2), 83-88.
- Telford, W. M., L. P. Geldart, R. E. Sheriff and D. A. Keys (1976). *Applied Geophysics*. Cambridge: Cambridge University Press.
- Wait, J. R. (1951). The Magnetic Dipole over the Horizontally Stratified Earth. *Canadian Journal of Physics* 29, 577-592.
- (1962). *Electromagnetic Waves in Stratified Media. [International Series in Electromagnetic Waves: 3]*. Oxford: Pergamon Press.
- (1969). Electromagnetic Fields of Sources in Lossy Media. In *Antenna Theory - Part 2*. ed. R. E. Collin and F. J. Zucher. New York: McGraw-Hill, pp. 468-471.
- (1971). Electromagnetic Induction Technique for Locating a Buried Source. *IEEE Transactions on Geoscience and Electronics* GE-9, 95-98.
- (1972). Locating an Oscillating Magnetic Dipole in the Earth. *Electronics Letters* 8(16), 404-406.
- (1982). *Geo-Electromagnetism*. New York: Academic Press.
- (1995a). What is "Quasi-Static". *IEEE Antennas and Propagation Magazine* 37(4), 109-110.
- (1995b). What is "Quasi-Static": Correction. *IEEE Antennas and Propagation Magazine* 37(5), 100.
- Walters, R. (1983). Subsurface Position Finding Using Dispersion Measurement Techniques (Final Year Student Project). *Department of Engineering*. York: University of York [unpublished; available from CREG library].

3 Specific Aperture: a Figure of Merit for Induction Loop Antennas

Abstract: A 'figure of merit' enables the performance of an induction loop antenna (transmitter or receiver) to be described in terms of 'cost factors' such as mass, material and size, without needing to define the design-specific quantities such as the number of turns and the wire diameter which, it is shown, do not affect its performance.

Before discussing transmitter and receiver antennas in detail, it is worthwhile formulating a figure of merit by which to rate them.

One of the fundamental tasks in designing a loop antenna – whether it be for transmission or reception – is how many turns of wire it should have. For a transmitter, we desire the amp-turns product to be as large as possible, but we also desire the power dissipation in the antenna to be as low as possible, which means that we have to consider the thickness of the wire as well. Clearly, it is possible to become embroiled in detailed design questions when a more fundamental view would be more helpful.

In this chapter, we will derive a figure of merit for loop antennas, and show how it can be used to characterise both transmitting and receiving antennas. The principles can be extended to electric dipole antennas and to radiating antennas.

It is convenient to refer to the figure of merit as the specific aperture, since its dimensions are $\text{m}^2/\sqrt{\Omega}$, although the term is, possibly, slightly misleading and should not be confused with the similarly named r.f. term, of effective aperture.

3.1 Deriving a 'Figure of Merit'

The source of the magnetic field of a loop of current is its magnetic dipole moment m_d . If the loop has area $A = \pi r^2$ (where r is the loop radius) and carries current I in N turns of wire, the magnetic moment is given by the familiar expression,

$$m_d = NIA \quad (3-1)$$

For an induction loop, the axial field strength at a distance $x \gg r$ is given, in terms of m_d , by the equally familiar expression – e.g. see (Ramo, Whinnery and Van Duzer, 1984) §2.10 – of,

$$H = \frac{m_d}{2\pi x^3} \quad (3-2)$$

which shows the inverse cube law of field strength v. distance. This remains valid provided that the quasi-static approximation, $k_0x \ll 1$, holds true; where k_0 is the wave number $2\pi/\lambda$. This will be discussed further in the chapter on propagation.

A small or low frequency induction loop, for which $k_0r \ll 1$, does not radiate any appreciable power, but clearly it can dissipate large amounts of power in its winding. Although the field strength is described by the parameter m_d , and therefore by the current I and the number of turns N , these are not fundamental cost factors. For design purposes, it would be helpful to describe the field strength H in terms of the mass of the antenna M and the power P dissipated in its winding.

In other words, when designing a radio system, the first task is to consider the size of the battery and the power drain. At that stage we are not interested in mundane aspects like how many turns of wire we should choose.

3.1.1 Demonstration of Principle

Consider an induction loop transmitter **A** comprising 100 turns of wire, carrying a current of 4A and having a resistance of 2Ω . We will compare it with a loop **B** of 400 turns, which has an identical mass.

Loop **B** will have four times the length of wire, so for the mass to remain the same as **A** it must be wound with wire of half the diameter of **A**'s. The resistance of loop **B** is now four times greater due to the length, and four times greater due to the smaller cross-sectional area so, overall, the resistance is 16 times greater, i.e. 32Ω . In addition, loop **B** has 4 times the turns so, for the same magnetic moment as **A** it needs only a quarter of the current, i.e. 1A. The power dissipation is I^2R , and for **A** this is $4^2 \times 2 = 32W$, whilst for **B** it is $1^2 \times 32 = 32W$.

Thus the two loops have the same mass, and use the same power; **A** is 400 turns \times 1A; **B** is 100 turns \times 4A and they both produce the same 400 Amp-turns of field.

Clearly, the performance of the transmitter loop – i.e. the magnetic moment in relation to the mass and diameter – is independent of the number of turns. We will now express this result more formally.

3.1.2 Specific Aperture

The magnetic field from an essentially non-radiating loop is produced at the expense of power dissipated within the physical resistance of the antenna winding. The resistance of N turns of wire of conductivity σ , cross-sectional area α , on a loop with perimeter p is

$$R = \frac{N p}{\sigma \alpha} \quad (3-3)$$

and the mass of the wire (of density ρ) is

$$M = N p \alpha \rho \quad (3-4)$$

Combining (3-3) and (3-4), we can write a figure of merit for the relationship between the power dissipation P and the magnetic moment m_d . Writing the figure of merit as Φ , we have

$$m_d = \Phi \sqrt{P} \quad \text{with} \quad \Phi = \frac{A}{p} \sqrt{M \frac{\sigma}{\rho}} \quad (3-5)$$

For a circular loop, we can replace the area / perimeter term with $\frac{1}{2} r$ to give

$$\Phi = \frac{1}{2} r \sqrt{M \frac{\sigma}{\rho}} \quad (3-6)$$

This shows that the relation between the power dissipation P and the magnetic moment m_d does not depend on the number of turns of wire, nor on its diameter, but only on the more fundamental properties such as the mass of the wire, its conductivity and density. Used at the design stage, this figure of merit shows that considerations such as the number of turns and the wire diameter are not of primary importance.

Although equation (3-6) defines a figure of merit, it does not provide an easy way to measure it for an existing antenna. However, using (3-1) we can write an alternative definition as

$$\Phi = \frac{NA}{\sqrt{R}} \quad (3-7)$$

which describes it using the number of turns, area and resistance of the antenna.

The figure of merit, Φ , expresses the fundamental parameter of *turns × area* relative to $\sqrt{resistance}$. If the *turns × area* product is considered to be an ‘aperture’ then the nomenclature of *specific aperture* is appropriate for Φ , for which the dimensions are $\text{m}^2/\sqrt{\Omega}$. The quantity \sqrt{R} represents the power dissipation in the transmitting antenna (actually \sqrt{power}) or the thermal noise of a receiving antenna and so *specific aperture* measures the performance of the antenna relative to these parameters. This figure of merit is essentially a theoretical result, but it does have practical applications, as will be discussed below; and it can also be used to compare antenna designs.

It is important to note that in (3-7) the resistance R is measured at the operating frequency, and not at zero frequency. Similarly, in (3-6) the conductivity must be adjusted to allow for the skin and proximity effects in the winding. Other obvious limitations apply – for example, the formula assumes that the turns of wire are identical.

3.1.2.1 Prior Art

This figure of merit, referred to as ‘effective aperture’ was first used by the author in 1990. However, the nomenclature was a little unfortunate because it could be confused with an unrelated term used at radio frequencies; and dimensionally, it is not an ‘aperture’. For this thesis the term ‘specific aperture’ has been used instead.

Other authors have derived similar figures of merit. (Burrows, 1978), §4.2, derives the power dissipation as

$$P = \frac{2^{8/3}(\pi^2 v)^{2/3} \rho^{5/3}}{\sigma M^{5/3}} m_d^2 \quad (3-8)$$

where v is the dimensionless ratio of the winding cross-section to the antenna area πr^2 . From this we could deduce the specific aperture to be defined by

$$\Phi = \sqrt{\frac{\sigma M^{5/3}}{2^{8/3}(\pi^2 v)^{2/3} \rho^{5/3}}} \quad (3-9)$$

There does not seem to be any particular advantage in expressing Φ in terms of v – in fact it would seem to make the expression needlessly complicated. We consider that, as a figure of merit, it is more instructive to express Φ using the loop radius r .

Unknown to the author at the time we developed our expression, American cave radio enthusiast Julian Coward (Coward, 1986) had also described what he called a ‘factor of merit’ which he defined, for a transmitter, as

$$\frac{1}{16\pi^3} \frac{\sigma}{\rho} \quad (3-10)$$

with a different constant appearing in his expression for a receiver. (Further comment on the use of the figure of merit with a receiving antenna will be made below). Coward made some general observations, including the main one made independently in by the author in 1990, that the number of turns was not significant.

3.2 Observations using ‘Specific Aperture’

3.2.1 Typical Values

Table 3-1, below, shows some typical values of specific aperture for an air-cored circular loop made of copper wire. These are the theoretical d.c. values and, as noted above, the a.c. performance will be different. All the loops listed in the table have a mass of 175g of copper. We can see that number of turns and wire diameter can be used to control the resistance of the winding whilst leaving the specific aperture unaltered. The largest of these antennas will have

ten times the range (transmitting or receiving) of the smallest, or will require 20dB less power for the same transmitting range. The skin depth in copper is 0.2mm at 100kHz. A wire diameter of 0.5mm is equivalent to 7/0.2mm equipment wire. Larger effective wire diameters than this will dictate special requirements (e.g. many thin turns, tape or braid construction), in order to minimise skin and proximity effects.

Loop diam. [mm]	Number of turns	Wire diam. [mm]	Specific aperture [m ² /Ω]	Resistance [Ω]
4000	22	0.3	33.8	66.8
2000	44	0.3	16.9	66.8
2000	16	0.5	17.0	8.7
1000	32	0.5	8.5	8.7
500	64	0.5	4.3	8.7
400	80	0.5	3.4	8.7
1000	8	1	8.5	0.55
500	16	1	4.3	0.55
400	20	1	3.4	0.55

Table 3-1 : Typical parameters for air-cored induction loops.
These loops would be suitable for cave radio applications.
All the loops in this table have a mass of 175g copper.

3.2.2 Field Strength is not Solely Related to Power

One important observation from (3-6) is that the field strength is not solely related to the power. This is entirely unlike the situation for a radiating antenna. This means that the regulatory specifications such as MPT 1337, (RA, 1988), which limit the power to an induction loop that is of unspecified ‘figure of merit’, are of doubtful worth. There is little point in limiting the power to 10W when the actual field strength that results from this depends on the mass and radius of the antenna. The new specifications such as ETS 300 330 (ETSI, 1994) are more sensible in that they place a direct limit on the magnetic field strength at a specified distance.

3.2.3 Number of Turns has no Effect on Performance

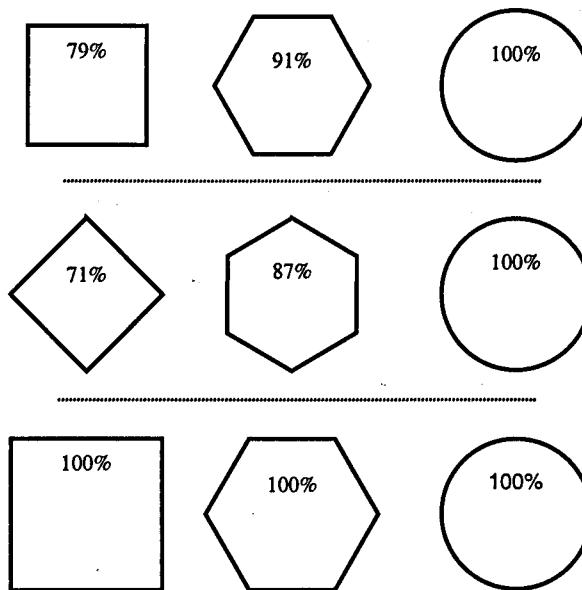
This fundamental conclusion simplifies the design and allows us to choose turns for other reasons, such as noise matching, impedance (power) matching, reduced skin and proximity effect, self-capacitance, tuning, available wire size etc. The fact that N is not an issue in the performance of the antenna was obvious from the example given at the start of this chapter.

3.2.4 Shape of Antenna

Loop antennas are often constructed to be collapsible, for easier transport. For operation, a large antenna can be spread out on the ground or, more often, attached to a collapsible frame. We can ask what effect the shape of the antenna has on its performance, given a certain length of the winding.

The specific aperture is proportional to the area of the loop divided by the perimeter (3-5) so, with a given amount of cable to form the winding (i.e. given mass and length), we find that when we have stretched the cable over its frame the specific aperture is proportional to the area. If the perimeter is p and the loop is circular then its area is $p^2/4\pi$. If it were square, its area would be $p^2/16$, so the square loop has a specific aperture that is $\pi/4$ time (79%) that of the circular loop. (See *Figure 3-1*, below). The specific aperture of an N -sided polygon, relative to a circular loop of the same mass and perimeter, can be shown to be

$$\frac{\pi}{N \tan(\pi/N)} \quad (3-11)$$



*Figure 3-1 : Specific Aperture of different shaped antenna loops.
When comparing antennas of equal mass, the shape has an effect on the performance.
Top row: Antennas of same mass & perimeter – a circular loop is preferred.
Middle row: Antennas of same mass and 'largest diameter' – circular loop is preferred.
Bottom row: Antennas of same mass and 'smallest diameter' – performances are equal.*

Instead of making the comparison for a given length of winding cable, we can compare two antennas of equal mass and equal 'largest diameter'. A circular loop has diameter $2r$ and so $A/p = \pi r^2 / (2\pi r) = r/2$. A square loop with a diagonal of $2r$ would have a side of length $\sqrt{2} \times r$ so its A/p ratio would be $r/(2\sqrt{2})$ showing that the square loop has a specific aperture that is $1/\sqrt{2}$ times (71%)

that of the circular loop. The specific aperture of an N -sided polygon, relative to a circular loop of the same mass and largest diameter, can be shown to be

$$\cos(\pi/N) \quad (3-12)$$

Similar results can be obtained for elliptical loops; and it can be shown that if the comparison is for ‘smallest diameter’, then A/p is always the same.

3.2.5 Aluminium is Better than Copper

The presence of the term σ/ρ shows that aluminium, with its lower density, is preferable to copper, even though its conductivity is not so high. For the same mass and power dissipation, an aluminium antenna will produce a 3dB increase in field strength. *Table 3-2*, below, shows the relative merits of some different materials.

Material	Density, ρ [10^3kg/m^3]	Conductivity, σ , [10^6S/m]	Merit $\sqrt{(\sigma/\rho)}$ [m/kg $\sqrt{\Omega}$]	Relative Merit [dB]
Sodium	0.97	21.3	148	5.3
Aluminium	2.7	37.0	117	3.2
Magnesium	1.74	21.7	112	2.8
Copper	8.93	58.1	81	0.0
Silver	10.5	62.5	77	-0.4
Gold	19.3	43.5	47	-4.6
Brass (typ.)	8.5	16.7	44	-5.2
Iron	7.86	9.5	35	-7.3

*Table 3-2 : The merits of different antenna materials.
Aluminium and copper are the only two viable materials for most applications,
the other possibilities all being expensive or reactive.*

The fact that the performance depends on the antenna material is not a new result. A similar result has already found application where it is required to minimise the power dissipation in a wire of a given mass and length. This is the case where a pylon has to support a power line, or a satellite has to support a woven wire tether (Chown, 1998) as well as applying to a short whip antenna (although the equation requires modification if the current distribution is not uniform). This can be seen in the example in §3.2.5.1, which follows.

3.2.5.1 Application to E-field Antennas

The resistance of a length ℓ of wire of conductivity σ and cross-sectional area α is

$$R = \frac{\ell}{\sigma\alpha} \quad (3-13)$$

and the mass of the wire (of density ρ) is

$$M = \ell \alpha \rho \quad (3-14)$$

If the current in the wire is uniform we can derive, in a similar exercise to that for specific aperture, the electric moment p_d (analogous to the magnetic dipole moment m_d) to be

$$p_d = I \ell = \sqrt{M \frac{\sigma}{\rho} \sqrt{P}} \quad (3-15)$$

which describes the power dissipation necessary to maintain a current I in a wire of length ℓ and mass M . This figure of merit could be described as ‘specific length’ but it is not as useful a figure of merit as specific aperture. For example, for a short wire antenna the main constraint on the dipole moment we can achieve may be the voltage stress on the antenna rather than its power dissipation. However, for long wire antennas ‘specific length’ may be a useful design parameter.

In a similar manner to (3-7), we can write this figure of merit in terms of the length and resistance of the antenna, as ℓ/\sqrt{R} . If the current in the antenna is non-uniform then ℓ is replaced by ℓ_e , the ‘effective length’ of the antenna. The quantity ℓ_e^2/R was described as a ‘figure of merit’ in (Burrows, 1978), §4.1.2.1. Note that effective length (or effective height) is a recognised r.f. term and has the dimensions of length. We will return to this, briefly in §3.4.2.

3.2.6 Q-Factor does not Depend on Number of Turns

Winding more turns of wire onto a loop increases its Q -factor because the inductance L increases with N^2 but R increases only linearly with N . However, this procedure also increases the mass and, if we keep the mass constant by decreasing the diameter of the wire, a different picture emerges. By writing the Q factor using Φ we can show that it then depends only on mass and *not* on number of turns. Using standard formulas e.g. (Ramo *et al.*, 1984), §4.8) for the inductance of a circular loop,

$$L_{loop} = N^2 \cdot \mu_0 r \ln\left(\frac{8r}{w} - 2\right) \quad (3-16)$$

and a solenoid,

$$L_{sol} = N^2 \cdot \mu_0 \frac{\pi r^2}{l} \quad (3-17)$$

(where r is the radius of the loop or solenoid, w is the width of the wire and l is the length of the solenoid), we can use (3-7) to derive the ratio of inductance to resistance as

$$\frac{L}{R} = \Phi^2 \cdot \frac{K_Q}{r^3} \quad (3-18)$$

where K_Q is a constant defined for a loop or a solenoid, by

$$K_Q|_{loop} = \frac{\mu_0}{\pi^2} \ln\left(\frac{8r}{w} - 2\right), \quad K_Q|_{sol} = \frac{\mu_0}{\pi} \frac{1}{2m} \quad (3-19)$$

where m is the *shape factor* of the solenoid, $m = l/2r$.

The Q factor of a tuned antenna is $\omega L/R$ and the bandwidth is $R/2\pi L$, which is therefore proportional to r^3/Φ^2 for a loop. Using (3-6) we can write

$$B_{loop} \propto \frac{r}{M}, \quad B_{sol} \propto \frac{l}{M} \quad (3-20)$$

showing that the bandwidth is proportional to the mass and independent of the number of turns and wire diameter. It also confirms that for a wide bandwidth the antenna must be light or physically large.

This result is only applicable where the system bandwidth is governed by the inductance to resistance ratio of the loop. This is the case for many conventional speech systems where we need to ensure that the tuned antenna has a sufficiently wide bandwidth. However, speech radio can benefit from equalisation, which negates this result. Additionally, ultra wide-band antennas, such as would be used for channel-sounding, require a different analysis. These cases will be discussed in a later chapter.

3.3 Using 'Specific Aperture' with a Receiving Antenna

Surprisingly, perhaps, the same figure of merit can be used with a receiving loop antenna. The law of induction gives the induced voltage U in terms of the incident flux ϕ as

$$U = N \frac{d\phi}{dt} = NA\omega\mu H \quad (3-21)$$

As before, we can eliminate N ; we use (3-7) and can thereby express the induced voltage U using the figure of merit derived earlier, to give

$$U = \omega\mu H \cdot \Phi\sqrt{R} \quad (3-22)$$

At first sight this may not seem particularly useful, because U depends on R , and so we have not completely eliminated N . However, it will now be demonstrated that R can be eliminated when we consider the signal/noise ratio (SNR).

Coward's analysis (Coward, 1986) did not consider SNR and so, to eliminate R , he assumed that the receiver was impedance-matched to a parallel-tuned loop; deriving an expression for the power in the receiver as $P = U^2/4R$. This approach is misleading because we are not necessarily considering matched systems. We are dealing with low frequencies, where the circuitry and antenna dimensions are much smaller than a wavelength and near fields, where power radiation is negligible. Consequently, issues of power transfer and power reflection are

not relevant. However, matching for noise resistance does play a part, and this complicated topic will be discussed in a later chapter.

3.3.1 Thermal Noise

The SNR is, of course, a more important parameter than the signal strength. The thermal noise voltage of the loop is given by $\sqrt{4kTBR}$, where k is Boltzmann's constant, T is absolute temperature, B is the noise-equivalent bandwidth and R is the resistance of the loop. Therefore the voltage SNR is

$$\left| \frac{S}{N}_{\text{thermal}} \right| = \frac{\omega \mu H}{\sqrt{4kTB}} \cdot \Phi \quad (3-23)$$

which shows that the receiver performance, too, is independent of number of turns, but related to the mass and physical size of the antenna.

The sensitivity of an antenna is sometimes defined, e.g. (Paschal, 1988) §2.2, as the signal's flux density for unity SNR in a 1Hz bandwidth. This alternative figure of merit, which we shall denote by Φ_s , is therefore defined by

$$\Phi_s = \frac{\sqrt{4kT}}{\omega \Phi} [\text{T Hz}^{-1/2}] \quad (3-24)$$

Paschal additionally defines a sensitivity that is 'normalised' with respect to frequency, by replacing ω by 2π in the above expression and he observes (§3.5) the dependence on mass and radius without explicitly creating a 'figure of merit'. Paschal notes that the terminology is somewhat unfortunate, since a more sensitive system has a smaller sensitivity. We observe that the units of Φ_s are $\text{T Hz}^{-1/2}$ and that Φ_s is the thermal noise spectral density of the receiver loop referred back to the magnetic field. But calling it *magnetic field equivalent spectral noise density* is not much more informative than *sensitivity* even if it is linguistically more correct.

3.3.2 Other Sources of Noise

There are other sources of noise to consider. A fuller discussion of the noise-matching of the amplifier to the antenna will be given in chapter 6 and, for the time being, we will consider only atmospheric noise. This is often described by an 'atmospheric noise temperature' T_a arising because a power of kT_aB is delivered via the antenna's radiation resistance R_r . Such arguments assume a matched system, so the equivalent noise voltage is $\sqrt{4kT_aBR_r}$. Defining the noise temperature ratio as $F_a = T_a/T_0$, and provided that the noise temperature has been derived from measurements made with a magnetic antenna, we can write

- i) Signal voltage

$$\omega \mu H \cdot \Phi \sqrt{R}$$

ii) Thermal noise

$$\sqrt{4kT_0BR}$$

iii) Atmospheric noise

$$\sqrt{4kT_0BF_aR_r}$$

and so the overall SNR is

$$\frac{S}{N}_{\text{atmos.}}^{\text{thermal+}} = \frac{\omega\mu H}{\sqrt{4kTB}} \Phi \sqrt{1 + F_a \frac{R_r}{R}} \quad (3-25)$$

This is important because, as we will show later, the term R_r / R is proportional to Φ^2 and so (3-25) shows that, for a large enough antenna, the SNR is independent of Φ . This result – that there is a maximum useful size for a receiver antenna – is well known for r.f. systems; but here, we have derived it using the concept of ‘specific aperture’ and applied it to essentially non-radiating systems. Because of this, it is helpful to remove the dependence on radiation resistance – which appears because of the way we have defined atmospheric noise – which we will now proceed to do.

3.4 Using ‘Specific Aperture’ with a Radiating Antenna

3.4.1 Radiation Resistance

We have described, above, how Φ is a useful way of describing the ‘efficiency’ of a non-radiating antenna. However, it can also be applied to a radiating structure, especially a small magnetic dipole ($r \ll \lambda$) with a low radiation resistance. The radiated power can be shown – e.g. using formulas in (Ramo *et al.*, 1984) §4.12a or (Kraus, 1950) §6.8 – to be

$$P_r = m_d^2 \frac{Z_0 k_0^4}{6\pi} \quad (3-26)$$

Using the definition in (3-7), we can write the ratio of radiated power to dissipated power in the antenna, i.e. the ‘efficiency’, as

$$\frac{P_r}{P_d} = \frac{R_r}{R} = \Phi^2 \frac{Z_0 k_0^4}{6\pi} \quad (3-27)$$

and we can note that, for the range of values of Φ and k_0 that we will be considering, the ratio of R_r / R is less than 0.005%.

3.4.1.1 Application to SNR equation

Substituting (3-27) into (3-25), and lumping some parameters together as an ‘atmospheric noise term’ F_k , allows us to write the SNR of a receiver loop as

$$\frac{S}{N}_{\text{thermal+atmos.}} = \frac{\omega\mu H}{\sqrt{4kTB}} \Phi \sqrt{1 + (F_k \Phi)^2} \quad (3-28)$$

with

$$F_k = \sqrt{F_a \frac{Z_0 k_0^4}{6\pi}} \quad (3-29)$$

In chapter 6, we will add a term to (3-28) that represents the amplifier noise, and we will introduce the concept of noise matching. For the time being, we can note that it is clear from (3-28) that there are two distinct situations. If the antenna is very small (i.e. the specific aperture is small or there is little atmospheric noise), with

$$F_k \Phi < \frac{1}{2} \quad (3-30)$$

then the SNR approximates (to within 12% or 1dB) to

$$\frac{S}{N}_{\text{thermal+atmos.}} \approx \frac{\omega\mu H}{\sqrt{4kTB}} \Phi \quad (3-31)$$

but that if the antenna is very large, with

$$F_k \Phi > 2 \quad (3-32)$$

then the SNR approximates to

$$\frac{S}{N}_{\text{thermal+atmos.}} \approx \frac{\omega\mu H}{\sqrt{4kTB}} / F_k \quad (3-33)$$

which is independent of the specific aperture. This will be discussed further in chapter 6.

3.4.1.2 Efficiency of Radiating Antennas

Consider a loop operating at 185kHz in air, and with the assumption that there are no effects caused by the proximity to the ground. For the quasi-static approximation to hold, we must be much closer to the antenna than $\lambda/2\pi$, which is 258m. If, however, we are much further away than this then the far-field is predominant. Equation (3-27) shows that the antenna efficiency will be around $4.5 \times 10^{-9} \Phi^2$ so if Φ is, say, $100\text{m}^2/\sqrt{\Omega}$ (which is a large specific aperture) the efficiency is still only 0.0045%. If the power amplifier delivers 50W to the loop, the radiated power will be only 2.2mW. Of course, this is only true for a small antenna – equation (3-27) is about 2% in error when the total wire length of the loop is around $\lambda/3$. For a 100-turn loop this is when the diameter is about $\lambda/1000$, or 2.6m at 185kHz. The radiation resistance of a ‘large’ loop is derived in, for example, (Kraus, 1950) and (Wolff, 1966).

3.4.1.3 Radiation Resistance v. Number of Turns

We can observe, from (3-26) that R_r is proportional to m_d^2 and thus to the number of turns squared; and thus note in passing that this demonstrates that radiation resistance is clearly a different quantity to ‘real’ physical resistance.

3.4.2 Effective Aperture and Effective Height

These two terms are commonly used in discussions of antenna theory. The power received by an antenna can be described in terms of a collecting area known as the effective aperture¹. This may be associated with a physical aperture but even a linear wire antenna can be described in terms of a notional collecting area. If the electromagnetic field has a power density of P_d , and the receiving antenna has an effective aperture of A_e , then it will capture a power of $P_d A_e$. The effective area depends only on the wavelength of the radiation and the directivity or ‘gain’ so we cannot use it as a figure of merit in our application.

The *effective height* or *effective length* of an antenna, h_e , is used to describe the voltage that the antenna generates in an electric field. In a field of strength E [V/m] the voltage is simply $E h_e$. Paradoxically, it is the effective *height* of an induction loop antenna that is a more useful figure of merit than the effective *aperture*. In a magnetic field H , the voltage induced in a loop is $\omega\mu H N A$ and so, knowing that for radiation in free space, $E / H = Z_0$, this leads directly to the standard expression for the effective height of a loop antenna, which is

$$h_e = k_0 \cdot NA . \quad (3-34)$$

This gives the effective height of a ‘whip’ antenna that would be needed to give the same signal voltage as obtained with the loop – provided, of course, that the antennas are in a radiation field where the ratio of E / H is known². We can write the effective height in terms of specific aperture as

$$\begin{aligned} h_e &= k_0 \cdot \Phi \sqrt{R} \\ &= \frac{\omega\mu}{Z_0} \cdot \Phi \sqrt{R} \end{aligned} \quad (3-35)$$

and this allows us to write the received voltage in (3-22) as

$$U = h_e Z_0 H . \quad (3-36)$$

This last relationship may have appeared to rely on the wave impedance of the medium being Z_0 but this is not so, because the definition of h_e also depends on Z_0 (via k_0 in (3-34)) such that the impedance terms cancel. In other words, h_e is merely a convenient figure of merit, and so (3-36) clearly holds for induction systems as well as radiating systems. However, although this result suggests that effective height could be used as a figure of merit, it does not help us to

- 1 In earlier published articles, the author confusingly described the term referred to in this thesis as *specific aperture* by the name *effective aperture*.
- 2 Note that the effective height is not the same as the physical height (or length) of a whip antenna, because it depends on the current distribution. For a short whip, the distribution is triangular and so the effective height is half the true height.

deal with the signal to noise ratio, and it does not give us access to the more fundamental antenna parameters, such as mass, which is why specific aperture is a more useful parameter for us to use.

3.4.3 Buried Transmitter Loops

For a submerged loop of radius r inside a radome of radius a , (Wait, 1952; 1969) derived the dissipation in the surrounding medium as

$$P \approx m_d^2 \frac{(\omega_e \mu_e)^2 \sigma_e}{12\pi a}, \quad r \ll a \ll \delta_e \quad (3-37)$$

where the external medium has properties μ_e , σ_e , ϵ_e ; is a good conductor with $\omega_e \sigma_e / \epsilon_e \gg 1$; and has skin depth defined in the usual way:

$$\delta = \sqrt{\frac{2}{\omega \mu \sigma}} \quad (3-38)$$

Attributing these eddy current losses to a fictitious 'external resistance' R_e allows us to express the loss in terms of Φ as

$$\frac{R_e}{R} = \Phi^2 \frac{1}{3\sigma_e \pi a} \frac{1}{\delta_e^4} \quad (3-39)$$

This expression is interesting because it is notably similar to (3-27). Instead of incorporating the wave number $k_0 = 2\pi/\lambda_{air}$ it contains the term $1/\delta_e$ but δ_e is related to the wavelength in the conducting medium (as derived in chapter 2) by

$$\delta_e = \frac{\lambda}{2\pi} \quad (3-40)$$

A typical value for R_e / R at 100kHz, with $\sigma_e = 1\text{mS/m}$, $a = 2\text{m}$ and $\Phi = 100\text{m}^2/\sqrt{\Omega}$, is 0.001% so R_e , in this example, is 8% which is a significant proportion of the transmitter power. We can see that the external resistance is far more significant than the radiation resistance which was calculated in an example in §3.4.1.2.

3.5 Concluding Remarks

We have derived a figure of merit, termed 'specific aperture', that can be used to describe the performance of induction loop antennas in terms of 'cost factors' such as mass and physical size. We have demonstrated several results that arise from a consideration of this figure of merit – for example, the number of turns of wire on the antenna is not of primary importance.

We shall see, in the following chapters, that the concept of specific aperture is invaluable, because it considerably simplifies and clarifies the design process for both transmitter and

receiver antennas. Previous authors (Burrows, 1978; Coward, 1986; Paschal, 1988) have observed the existence of a possible figure of merit, but without taking full advantage of its potential.

3.6 References³

- Burrows, M. L. (1978). *ELF Communications Antennas. [IEE Electromagnetic Waves Series: 5]*. Stevenage: Peter Peregrinus.
- Chown, M. (1998). Knitting in Space. *New Scientist*, 29th August 1998, 43-44.
- Coward, J. (1986). Some Thoughts on Cave Radio Antenna Design. *Speleonics* 4, 6-7.
- ETSI (1994). *European Telecommunications Standard ETS 300 330*. Valbonne, France: ETSI Secretariat.
- Kraus, J. D. (1950). *Antennas*. New York: McGraw Hill.
- Paschal, E. W. (1988). *The Design of Broad-Band VLF Receivers with Air-Core Loop Antennas*. Anacortes, WA: Whistler Radio Services. (2nd Edition).
- RA (1988). *MPT 1337 Performance Specification*. London: Department of Trade & Industry, Radiocommunications Agency.
- Ramo, S., J. R. Whinnery and T. Van Duzer (1984). *Fields and Waves in Communication Electronics*. New York: John Wiley. (2nd Edition).
- Wait, J. R. (1952). The Magnetic Dipole Immersed in a Conducting Medium. *Proc. IRE* 40(10), 1244-1245.
- (1969). Electromagnetic Fields of Sources in Lossy Media. In *Antenna Theory - Part 2*. ed. R. E. Collin and F. J. Zucher. New York: McGraw-Hill, pp. 468-471.
- Wolff, E. A. (1966). *Antenna Analysis*. New York: John Wiley & Sons.

³ *Speleonics* is the newsletter of the Communications and Electronics Section of the USA-based National Speleological Association – see www.caves.org/section/comselect

4 Antenna Design

Abstract: *The classification of antenna systems into electric, magnetic and current sources is discussed, but the antenna of choice for portable communications is the induction loop, two basic forms of which are the air-cored loop and the ferrite-cored solenoid. A comparison is made between these two forms and it is concluded that a communications system may benefit from having a separate transmitter and receiver antenna, to different designs. It is demonstrated that no clear-cut decisions on antenna systems can be made without knowledge of the design specification for the system. Several 'exotic' antenna designs are also considered and would benefit from further study.*

The communications systems we are discussing rely, for a large part, on near-field effects. In this situation, unlike that of conventional radio, the behaviour of the fields cannot be analysed without considering the type of antenna in use. For example, the magnetic near-field from an electric dipole does not have the same spatial dependence as the magnetic near-field from a magnetic dipole and yet both produce (in air) the same propagating far-field.

In chapter 2, we discussed the fields from an induction loop transmitter, i.e. a small magnetic dipole. Such an antenna is convenient to analyse because it is small (i.e. it can be treated as a point source), it is isolated (i.e. it does not interact with its surroundings by stray capacitance) and it is a good example of the type of antenna we might want to use in a portable communications system. It is, however, not the only antenna that we could consider.

In this chapter we will give a brief summary of some other possible antenna systems, and we will then discuss some characteristics of induction loop antennas in detail. Then, in the following chapters, we will give information specific to transmitting and receiving antennas based on portable induction loops.

4.1 Antenna Classification

A loop of wire carrying a steady current is a source of magnetic flux. Similarly, a pair of electrodes or wires holding a steady charge is a source of electric flux, and a pair of electrodes buried in the earth is a source of current flux. This forms the basis for classifying antennas as *B*, *D*, or *J*-flux devices. Conventional nomenclature describes *B* and *D*-flux antennas by the field

they generate – H or E – so we will adopt this, adding J to describe the field from a current-flux antenna, which is convenient but not entirely consistent. (*Table 4-1*).

H :	Magnetic flux density	B [H m^{-2}]	is related to Magnetic field strength	H [A m^{-1}]	by $B = \mu H$ (μ is permeability)
E :	Electric flux density	D [F m^{-2}]	is related to	E [V m^{-1}]	by $D = \epsilon E$ (ϵ is permittivity)
J :	Current flux density	J [A m^{-2}]	Electric field strength		by $J = \sigma E$ (σ is conductivity)

Table 4-1 : H , E and J fields - relationship between field strength and flux density.

A magnetostatic source will generate a magnetic field that will penetrate a perfect electrical conductor, whereas an electrostatic source will generate an electric field that is terminated by charges on the surface of the conductor, with no penetration of the conductor by the field. Extending this observation to less-than-perfect conductors and low frequencies, we noted, in chapter 2, that the high degree of attenuation of the electric field, as it crossed the air/rock boundary, indicated that it was less suitable than the magnetic field for sub-surface communications.

- A time-varying magnetic field must co-exist with a time-varying electric (E) field so,
- although an E -field antenna may not produce a useable electric field underground, its magnetic field could be detected with an H -field antenna. Although we would not normally choose to receive the electric field, there may be circumstances, at higher frequencies, where it was feasible,. In addition, eddy currents will be generated in the conducting medium, giving rise to a J field, which we could detect with a pair of electrodes.

Extending this argument, each of the three transmitting systems – H , E , and J field– can be detected by an H , E , or J -field receiver, giving rise to three ‘pure’ antenna systems and six possible ‘hybrid’ systems to consider. Even for a ‘pure’ system there will be a considerable difference in operation between an air-cored induction loop and a ferrite-cored solenoid, or between a long wire ‘whip’ and a pair of dielectric-separated discs. After some brief discussion of antenna systems we will restrict our analysis to induction loop systems (air-core and ferrite-core). However, much time could be spent investigating novel antenna systems and this could be the subject for future work.

4.1.1 H-field Antennas

Induction loop transmitter and receiver antennas are the obvious (and traditional) choice for sub-surface operation. Two standard configurations are the air-cored loop and the long thin solenoid; the latter featuring a core of ferrite or other high permeability material. The H -field antenna is easy to analyse for the reasons given earlier, namely that it can be considered as a

point source, and it is usually isolated from its surroundings in the sense that there are no 'stray' effects that are difficult to model. (But clearly it operates, in part, by mutual inductance so it is not isolated in that sense).

4.1.2 E-field Antennas

The electric field transmitter antenna is more difficult to analyse than *H*-field devices and is usually not practical to use. A loop antenna requires a low voltage at a high current, and presents a high inductance, whereas an electric dipole requires a high voltage (perhaps in the region of kilovolts) at a low current and presents a low capacitance load to its driver, which makes modelling difficult, due to the significance of stray capacitance. This is not just a theoretical problem – the performance of a 'whip' antenna in the field will be dependent on its surroundings, and will not be entirely predictable. Used as a receiver, the antenna presents similar problems, with the low capacitance (perhaps only 10pF) presenting a very large reactance, which would be difficult to noise-match to the receiver amplifier¹, and prone to swamping by strays and leakage.

E-field antennas based on a vertical monopole, supported by a mast or balloon, and horizontal arrays of wires driven at multiple points have been used for VLF and ELF transmission in systems where atmospheric propagation of the far-field was required. The use of such antennas for subterranean work has not been a major feature of past work because, in general, an induction loop will give a better performance. However, an *E*-field antenna does generate a magnetic field and, in some circumstances, this field could be useable underground.

In principle, we can derive the fields from an electric dipole in the same way that we derived the fields from a magnetic dipole in chapter 2. However, it was assumed, for the magnetic dipole that the current distribution was uniform. We cannot make that assumption for a straight wire because the current distribution depends on the capacitance of the structure and the proximity to ground.

The current distribution in a non-resonant end-fed vertical monopole over a ground plane is triangular – i.e. it decreases linearly from a maximum at the feed-point to zero at the end of the antenna. The dipole moment of the antenna can be increased by forcing the current to the end of the wire, which is easily achieved by increasing the capacitance at this point by means of a horizontal wire, or mesh, to form a 'T' or inverted 'L' antenna². The dielectric disc antenna makes further use of this principle by having the antenna form two plates of a capacitor, which can be filled with a high permittivity material. Such a structure is the *E*-field equivalent of a

1 Noise-matching will be discussed in chapter 6.

2 A review of such antennas is given by (Belrose, 1992).

ferrite-cored solenoid. We will show, later, that a solenoid benefits from being as long and thin as possible. In a similar fashion, the dielectric disc antenna should be as wide and flat as possible to reduce fringe effects. Dielectric discs have been considered for use on the hulls of submarines, where the hull forms one plate of the capacitor. Few substances have a relative permittivity greater than ten, so the advantage of a dielectric disc is minimal in our application.

For a horizontal wire, the effect of the capacitance to ground can be severe. Modelling the wire as a transmission line, we can show that the attenuation of the current is exponential with distance. For this reason, large arrays of horizontal wires, for fixed ELF installations, can be driven at multiple points; and the wire suspended above the ground to reduce the capacitance. Another method of end-loading a horizontal wire is to bury the ends in the earth. The effective size of an earth electrode depends on the nature of the region surrounding it, but if there is a sufficiently large region of a high conductivity then the electrodes will have a significant mutual capacitance, thereby drawing the charge to the ends of the wire. The primary purpose of this, for radiating ELF systems, is to aid the generation of fields from the dipole current. However, for sub-surface use we can also consider the field generated by the current flow through the earth itself, and this may be the more significant field underground. This leads us to the concept of a current injection or *J*-field antenna, which we will discuss below.

4.1.2.1 Comparison of *E*- and *H*-field Antennas

Although the electric field suffers an attenuation at the air/rock interface, it would still be usable if we could generate a sufficiently large electric field at the antenna – or if an *E*-field antenna generated a sufficiently large magnetic field. Therefore, to complete this discussion we need, briefly, to compare the two antenna types. We can use the concept of specific aperture, developed in chapter 3, to describe the magnetic dipole moment of a induction loop antenna; and we can use the related figure of merit – specific length – to describe the electric dipole moment of an electrostatic or wire antenna. The signal to noise ratio of a receiver antenna can also be expressed using these figures of merit, although the detail of this is not given until chapter 6. The relevant formulas are listed in *Table 4-2* below. Where the field expressions are not given in chapter 2, they are readily available in textbooks. See, for example, (Clemmow, 1973).

The nomenclature in *Table 4-2* denotes the specific aperture (or length) of the antenna by Φ , with a superscript denoting transmitter or receiver, and a subscript denoting electric or magnetic dipole. If the transmitter or receiver antennas make use of a magnetic core or a dielectric then μ , and ϵ , need to be added to the expressions as appropriate. But we should note that the expression for the electric field attenuation does *not* depend on the permittivity of the sub-surface media. The collection of terms that comprise each formula are

- a) the dipole moment as a function of the antenna power dissipation
- b) the path attenuation
- c) where applicable, the electric field attenuation
- d) the induced signal at the receiver, expressed as a function of $\Phi\sqrt{R}$,

and the thermal noise term, $\sqrt{(4kTBR)}$, with the R cancelling that in $\Phi\sqrt{R}$.

	Magnetic dipole transmitter	Electric dipole transmitter
Magnetic dipole receiver	$\Phi_m' \sqrt{P} \frac{1}{4\pi r^3} \frac{\omega\mu_0 \Phi_m'}{\sqrt{4kTB}} \quad \{\Phi_m' \Phi_m'\}$	$\Phi_e' \sqrt{P} \frac{1}{4\pi r^2} \frac{\omega\mu_0 \Phi_m'}{\sqrt{4kTB}} \quad \{\Phi_e' \Phi_m' r\}$
Electric dipole receiver	$\Phi_m' \sqrt{P} \frac{\omega\mu_0}{4\pi r^2} \frac{\omega\epsilon_0}{\sigma} \frac{\Phi_e'}{\sqrt{4kTB}} \quad \left\{ \frac{\Phi_m' \Phi_e' r}{\sigma/\omega\epsilon_0} \right\}$	$\frac{\Phi_e' \sqrt{P}}{\omega\epsilon_0} \frac{1}{4\pi r^3} \frac{\omega\epsilon_0}{\sigma} \frac{\Phi_e'}{\sqrt{4kTB}} \quad \left\{ \frac{\Phi_e' \Phi_e' / k_0^2}{\sigma/\omega\epsilon_0} \right\}$

Table 4-2 : Comparison of electric and magnetic dipole systems.

The expressions show the receiver signal / noise ratio as a function of the transmitter and receiver antenna types, with symbols as defined in the main text. The expressions in curly brackets show the relative SNR, after common factors have been removed from the main expressions.

Considering a magnetic field receiver, we can see from the table that, for the same transmitter power, the electric dipole transmitter confers an advantage over the magnetic dipole transmitter of

$$\frac{\Phi_e'}{\Phi_m'} r \quad (4-1)$$

which suggests that it can be used as a source of a viable magnetic field in a configuration we will denote by $E \rightarrow H$ (shorthand for an *E-dipole* transmitter sending to an *H-dipole* receiver).

A typical magnetic transmitter antenna with 500g of copper and a linear dimension of 1m will have $\Phi_m \approx 14\text{m}^2/\sqrt{\Omega}$. The comparable specific length for an electric dipole operating at $r = 20\text{m}$ would therefore be $\Phi_e \approx 0.7\text{m}/\sqrt{\Omega}$. This antenna would have a very low mass ($75\mu\text{g}$) and we can surmise, without calculation, that dissipating 10W in such a low mass of wire will require a difficult design. The conclusion is that the concept of specific length is not as useful as that of specific aperture. We can support this inference by noting that the magnetic moment from an electric dipole is simply $m_d = \Phi_e' \sqrt{P}$ which suggests that a static electric dipole will generate a magnetic field. But this is clearly false; the flaw being that to dissipate a power P in the static antenna would require an infinite driving voltage. This situation arises because we have chosen to describe the magnetic field in terms of an electric *current* dipole moment. If the

magnetic field is written in terms of the *electrostatic* dipole moment (with $I = j\omega q$) then the expected frequency dependence can be seen.

Reasoning from more practical parameters, we can see that, for the antennas to produce the same magnetic field at 20m, we would require an electric dipole moment of 2.2Vm. If the antenna were constructed from two metal plates, each 100mm square, and separated by 10mm of air, we would need to maintain 220V across the plates, and counteract a capacitance of 8.85pF. Clearly, the electric dipole moment is best described in terms of the antenna voltage U and its capacitance C as

$$p_m = j\omega CU \cdot \ell \quad (4-2)$$

rather than as $p_m = \Phi'_e \sqrt{P}$, and it is the large voltage requirement and low capacitance that makes this type of antenna difficult to use. However, we can see that a long wire – even if it is not earthed at the ends to form a *J*-field antenna – could be viable in certain situations.

Further inspection of *Table 4-2* reveals that no other combination of antennas is feasible. However, noting that

$$\frac{1}{k_0^2} \frac{\omega \epsilon_0}{\sigma} \equiv \frac{1}{2} \delta^2 \quad (4-3)$$

we might suppose that an $E \rightarrow E$ system would work at low frequencies where the skin depth was large. However, this deduction is flawed for the reason noted earlier – that it requires an unfeasibly large antenna voltage (tending to infinity at zero frequency).

4.1.2.2 Radiating Structures

We have seen that for most practical forms of small portable loop antenna the radiation resistance is negligible. However, it is possible that, in very dry rock, we could wish to use a frequency as high as 1MHz, for which a small multi-turn loop would have a noticeable radiation resistance. Provided that the total length of wire in the loop (i.e. $2\pi r N$) remains less than about $\lambda/3$ the radiation resistance is accounted for in the analysis of §3.4. However, our ‘figure of merit’ – specific aperture – needs to be modified to take into account the power dissipated in R , and so we write

$$m_d = \Phi \sqrt{P} \cdot \frac{1}{\sqrt{1 + R_r/R}} \quad (4-4)$$

where the basic specific aperture Φ is still defined in terms of the ‘ohmic’ resistance of the loop, as in (3-7). The quantity R_r/R is usually negligible. It was defined in (3-27) in terms of Φ , and so we can see that there is a limit on the specific aperture because, if Φ becomes so large that $R_r \gg R$, we have

$$m_d = \sqrt{\frac{6\pi}{Z_0 k_0^4}} \cdot \sqrt{P} \quad (4-5)$$

which is independent of Φ . For example, at 1MHz, however large we make Φ its net effect will be limited to $510\text{m}^2 / \sqrt{\Omega}$. This equation is, however, of limited application, because it only applies if the loop is smaller than a self-resonating structure; and to achieve $R_r \gg R$ with such a loop would be difficult, with factors such as inter-winding capacitance and PA dissipation serving to reduce the efficiency.

A single-turn loop with a perimeter of a wavelength is self-resonant and at 1MHz this would be the case for a loop of diameter of 95m; for comparison an electric half-wave dipole would have a length of 150m. We can infer from the presence of a radiation resistance that the antenna will radiate power and this may not be desirable if this means that significant power is lost to the air. However, the radiation resistance is affected considerably by the surface impedance of the ground, η , (see equation (2-21)) reducing to $R_r \eta / Z_0$ for the case of a horizontal electric dipole with an ‘intrinsic’ or free-space radiation resistance of R_r . However, directive transmission into the ground may be possible (Smith and An, 1983).

Loops that have a perimeter larger than about $\lambda/3$ have radiation properties that are more difficult to describe than those of a small loop – see for example, discussion in (Balanis, 1997; Kraus, 1950; Wolff, 1966). Another factor to consider in the analysis of a large antenna is that it cannot necessarily be treated as a point source for the communications distances that we need to consider.

Although large antennas do have a place – particularly for communications in the mining industry – we will not discuss them further in this thesis, in which we are concentrating on the use of small, portable antennas.

4.1.3 J-field Antennas

Here we will consider systems where the current injection into the ground gives rise to a current flow that could be envisaged as forming part of the antenna. However, current injection systems need not involve ‘antenna’ principles at all – the telegraphs of the 19th century used a single wire and an earth return and therefore depended on the transmission of current through the ground. Earth-return, or *single wire telephones* (SWT) are important for cave communications and will be briefly discussed in §4.1.4 for completeness.

For near-field purposes, we can model the antenna as an electrode capacitance C in parallel with the ground resistance R ; and we will assume that we can ignore the capacitance of the feed

wire to ground³ (*Figure 4-1*). The proportion of the current that flows through the ground depends on the ratio of C to R . We assume that there is no contact resistance, and that the electrodes are half-buried spherical shells. The resistance between two well-spaced such electrodes of radius a , in a medium of conductivity σ is

$$R = \frac{1}{\pi \sigma a} \quad (4-6)$$

The capacitance is given by a similar formula to the *conductance*, but we have to allow for the field in air, as well as the field in the ground, so we have

$$C = \pi \epsilon_0 a (1 + \epsilon_r) \quad (4-7)$$

where ϵ_0 is the permittivity of the air path, and ϵ_r is the relative permittivity of the ground⁴. Assuming, for simplicity, that $\epsilon_r \gg 1$, which is good enough for many situations, the ratio of the current in the feed wire to current in the ground is found, from simple lumped-circuit theory, to be

$$1 + j \frac{\epsilon \omega}{\sigma} \quad (4-8)$$

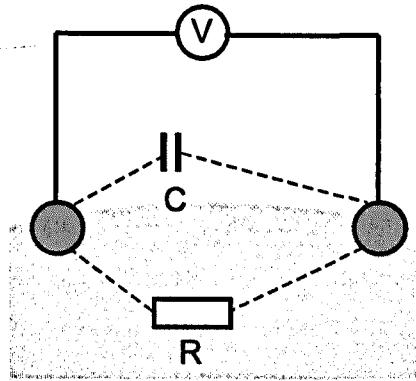


Figure 4-1 : A simple model of a J-field 'earth-current' antenna.
Ignoring stray capacitance around the feed wire, we can model this antenna with a capacitance (due to the earth electrodes) in parallel with the earth resistance.

Note that this result does not depend on the size of the electrodes, because R and C are equally effected by the electrode size. However, for the formula to be useful, the electrode capacitance must be much greater than the capacitance of the feed wire – i.e. either the

- 3 Using the standard formula for the capacitance of a twin-wire line for wires of radius a and separation $2b$, with $a \ll b$, and assuming the second wire to be a reflection of the first in the perfectly conducting earth, we can arrive at a figure of 5pF/m . For this we assume $a = 0.25\text{mm}$, $b = 25\text{mm}$ (allowing for grass and small rocks) and we use $C \approx \pi \epsilon_0 / \ln(2b/a)$. Increasing b does not decrease the capacitance much, but a lower b causes it to increase significantly. Using an exact formula – e.g. (Clemmow, 1973) §3.3.3 – and with $b = 2a$, we have $C = 21\text{pF/m}$.
- 4 If we assume $a = 200\text{mm}$ and $\epsilon_r = 9$ then we have $C \approx 56\text{pF}$ which is ten times the capacitance per unit length of the feed wire. Clearly we can make the electrode capacitance larger by using larger electrodes and by ‘watering’ the region around them. We can also note that the hemisphere of radius 200mm has a similar resistance and capacitance to that of a thin rod of length 1m . (Gibson, 1997d).

capacitance of the feed-wire to ground, if the ground is a good conductor, or the capacitance between its arms, if the ground is a poor conductor.

The crucial term in this equation is $\epsilon\omega/\sigma$ which defines the *characteristic frequency*. If the rock is a ‘poor’ conductor, with $\epsilon\omega/\sigma \gg 1$ then – in our lumped circuit model of *Figure 4-1* – most of the current flows in the *C* arm of the circuit, and we have an end-loaded electric dipole. Conversely, if the rock is a ‘good’ conductor – which is the case for the systems we are considering – then most of the current flows in the *R* arm and our antenna will have some of the characteristics of a large loop. The effective area of the current loop can be shown (Burrows, 1978) §3.3, to be $\ell\delta\sqrt{2}$ where ℓ is the distance between the earth electrodes and δ is the skin depth in the ground, which is assumed to be uniform.

An analysis of this antenna is not straightforward. Firstly, there is a field contribution from the proportion of the current that flows only in the electric dipole. Secondly, there is the field due to the current in the earth to consider, which does not result from a single, simple, path. In chapter 2, we saw that we could – in principle – integrate the current flow along a loop to obtain the magnetic vector potential, but this would be a daunting task for this distributed earth current. Fortunately, this is not necessary because, assuming an isotropic medium, the *J*-flux in the ground is coincident with the *D*-flux from the electric dipole antenna, and thus we can expect that the field could, with suitable boundary conditions, be derived from an analysis of the electric dipole antenna alone. One conclusion we can draw from this is that the shape of the magnetic field lines is one of concentric rings around the antenna. An analysis of the fields from a grounded wire was undertaken by (Hill and Wait, 1973). Recently, (Lippold, 2000) has described how the equations can be modelled on a computer, using MathCad.

4.1.3.1 *The HeyPhone*

It is worth mentioning one particular radio system here – the HeyPhone, designed by UK radio amateur John Hey. This consists of a single-sideband radio operating at 87.5kHz into a tuned loop or a pair of earth electrodes. It was developed with the help of the British Cave Research Association’s Cave Radio Group and the British Cave Rescue Council. One of the aims of the project was to place the design information in the public domain – it is online at [caves.org.uk / radio / heyphone](http://caves.org.uk/radio/heyphone).

The salient point about the HeyPhone is that its preferred mode of operation uses earth electrodes at either or both ends of the communications link. The use of earth electrodes at the transmitter has resulted in a considerable boost to the communications range and the ease of operation.

4.1.4 Cable-based Communications

Although the scope of this thesis is ‘wireless’ communications, there are circumstances where cable-based systems are appropriate or even preferable. Some cable-based systems can be described as a hybrid with a wireless system and so it is of value to briefly review them, although we shall not consider them further in this thesis.

4.1.4.1 Two wire telephones

These need no further comment other than to say that they are hardly ever used, because there is a considerable advantage in using a so-called *single-wire telephone*.

4.1.4.2 Single-wire telephones

Traditionally, these used to be known as ‘earth return’ telephones, but we have promoted the use of the term ‘single wire’ because it is not always certain that the return path is through the ground. An audio base-band device works on the principle that the conductivity between two earth electrodes does not depend on the distance between them (Gibson, 1997d). This principle was used in the telegraphs of the 19th century. High-impedance SWTs have been developed and we have shown that these can cope with a short to ground midway along the wire (Gibson, 1997b). We have also shown that a high-impedance device does not need a ground connection at all – there being sufficient capacitance between the human body and earth for a.c. audio signals to be transmitted, even if there is no d.c. path (Gibson, 1997c).

If the audio is modulated onto an LF carrier than it becomes possible to tap into the telephone line using an inductive or capacitive coupler. Such devices are in use by the mining industry and the fire service as well as in cave communications. One such LF carrier system used by the mining industry uses a bare steel hawser as the signal wire (Gibson, 1997e; 1998). The signal attenuation along such a wire, if it was in continual contact with the ground, would be exponential with distance.

4.1.4.3 Hybrid with Induction Radio

A signal transmitted along a long telephone wire, with an earth return path, forms a long thin induction loop and so the signal – especially if impressed upon an LF carrier – can be detected by an induction loop. Such a system is almost identical to the HeyPhone J-field device described above. With these systems the exact mode of propagation of the signal is not always easy to define, and so the analysis can be rather difficult.

4.1.4.4 Guide-Wire Radio & Leaky Feeders

From the LF-carrier SWT it is a small step to considering ‘guide-wire’ radio where the wire, in some sense, guides or re-radiates an RF signal in the HF or VHF bands.

4.2 Choice of Antenna System

We have seen that there are three basic types of antenna, and that these can be used in a number of combinations. Conventionally, portable systems for cavers and miners have made use of air-cored induction loop antennas, but ferrite-cored solenoids have also been tried, as have large fixed loop installations. For large systems the choice extends to more expensive installations such as masts, the line antenna arrays used for VLF submarine communication, and large, towed antennas.

For our analysis of through-rock communications we will limit our study to that of portable induction loop antennas of various designs. Although recent experiments by cavers have demonstrated that earth-current antennas give stronger signals, induction loops are more portable and versatile; and so, if a detailed analysis can tell us how to improve such a system, it will be preferable to using large earth-current antennas.

However, the choice of antenna system does not rest here. Even if we have decided to use an induction loop there are still many design questions still to answer. We saw in chapter 3 that a small heavy antenna could have the same performance as a larger, less massive device, but this does not address the question of bandwidth. We also have to make a choice between an air-cored loop and a ferrite-cored solenoid, and this choice may depend on the application. In the following sections we will look at induction loop antennas in more detail.

4.3 Designing Induction Loops using Specific Aperture

In chapter 3 we showed that we could write a figure of merit for the relationship between the power dissipation P and the magnetic moment m_d . Writing the figure of merit as Φ , we defined

$$m_d = \Phi \sqrt{P} \quad \text{with} \quad \Phi = \frac{1}{2} r \sqrt{M \frac{\sigma}{\rho}} \quad (4-9)$$

where r is the radius of a circular loop and σ, ρ are the electrical conductivity and mass density of the wire used in the winding.

This important result shows that the relation between the power dissipation P and the magnetic moment m_d does not depend on the number of turns of wire, nor on its diameter, but only on the more fundamental properties such as the mass of the wire, its conductivity and density. Used at the design stage, this figure of merit shows that considerations such as the number of turns and the wire diameter are not of primary importance.

4.3.1 Application to Ferrite-cored Antennas

The expression for specific aperture, quoted above, applies to an air-cored loop. For a magnetic-cored antenna, the magnetic moment is increased by the effective relative permeability of the material μ' , and we can write

$$m_d = \mu' \Phi \sqrt{P} \quad (4-10)$$

and for convenience we define a '*magnetic*' specific aperture (MSA) as

$$\Phi_m = \mu' \Phi \quad (4-11)$$

The mass term that appears in the expression for the specific aperture refers to the contribution from the winding alone. For a ferrite-cored antenna most of the contribution to the mass will be from the core so we need to develop a figure of merit that includes this parameter. Obtaining a succinct expression is difficult because of the number of parameters involved, and the change in significance. For example, mass and radius are a concise way to describe an air-cored loop, but length and radius are more appropriate for a ferrite-cored solenoid. For the time being we will limit ourselves to describing a relationship between the mass of the winding M_w , that is used in the expression for specific aperture, and the mass of the core, M_c .

We will observe, in §4.4.2.2, that the winding on a ferrite rod should be concentrated over the central section. With parameters as defined in *Figure 4-2* we can write the mass ratio as

$$\frac{M_w}{M_c} = 4 \frac{t}{2r} \frac{\ell_w}{\ell_c} \frac{\rho_w}{\rho_c}, \quad \text{with } t \ll 2r \quad (4-12)$$

where, for a particular design, we might expect each of the three ratios on the right-hand side of the expression to remain constant. For example, we might have $r = 4\text{mm}$, $t = 1\text{mm}$, so $t/2r = 1/8$, $\ell_w/\ell_c = 1/3$, and $\rho_w/\rho_c = 8930/4800$ for a small solenoid with a copper winding⁵, so the ratio of winding mass to core mass is about 0.31 for this class of antennas.

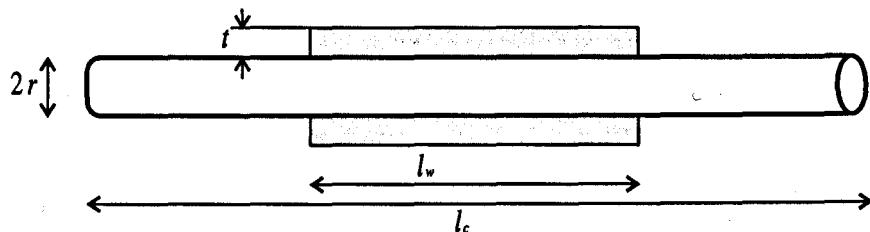


Figure 4-2 : Ferrite rod antenna with winding on central section.

We have seen that the specific aperture is proportional to the product of r and $\sqrt{M_w}$. For an air-cored loop the radius of each turn of wire will be about the same, but for a ferrite rod this is

⁵ A typical ferrite material, type 3C85, has a mass density of 4800kg m^{-3} . The figure for copper is 8930kg m^{-3} .

not so. We can show that, for a ferrite rod, the *area / perimeter* term in Φ (see equation 3-5) is replaced by *area of ferrite core / mean perimeter of winding* and that this results in an ‘effective radius’ for the winding of

$$r_{\text{eff}} = r \left(1 + \frac{t}{2r} \right) \quad (4-13)$$

which we must use in any calculations of Φ that are based on mass and radius. We can observe that the effective radius is smaller than the radius of the ferrite rod. An alternative method of analysis would be to reduce the effective permeability of the core, to allow for the fact that it does not entirely fill the cross-sectional area of the turns on the winding.

4.3.2 Further Observations using Specific Aperture

In chapter 3 we saw that the concept of specific aperture allowed us to express several useful results for induction loop antennas that would aid us during the design process. These included

- i) The field strength is not solely related to power input to the antenna.
- ii) The number of turns on the antenna has no effect on the performance.
- iii) Aluminium may be a better material for the winding than copper.
- iv) The Q -factor of the antenna does not depend on number of turns.

We will now briefly mention some additional observations.

4.3.2.1 Sixth-power Law for Range

It was sometimes stated by early cave radio experimenters that the power conveyed in an induction loop system followed the inverse sixth power of distance. Since the induced voltage follows $1/r^3$ there is some truth in this, but since the induction field does not actually radiate power this is a slightly misleading statement to make. The, perhaps erroneous, statement that “the power drops off as $1/r^6$ ” was sometimes given as the reason why cave radios did not cause long distance radio interference. In reality, the situation is slightly different. The elements of the field that give rise to long distance radiated interference drop off as $1/r$, but they are at a very low level because the loop is such an inefficient radiator.

However, since the signal strength is proportional to the magnetic moment, and therefore to the square root of transmitter power – see (4-9) – then it is true to say that the *range* is proportional to the sixth power of the power dissipation in the transmitting antenna. That is,

$$D_{\max}^6 \propto \Phi^2 P \quad (4-14)$$

where D_{\max} is the range.

This has obvious implications for system design – to double the range requires a 64-fold increase in transmitter power, or merely a four-fold increase in both radius and mass of the antenna – assuming operation remains mainly within the near-field.

4.3.2.2 *Third-power Law for Range*

An adaptive communications system could make use of the sixth-power law, described above, to ensure that maximum transmitter power was only used when necessary. However, we will show, in the next chapter, that the efficiency of an analogue power amplifier (PA) depends on the output voltage and, as we will now see, this causes a change in the power-law from sixth order to third order.

Suppose our PA were set up with a maximum gain, to produce a transmitter loop power of 64W and that the PA was near to 100% efficiency. Now suppose that we desire to halve the transmission range by reducing the loop power 64-fold to 1W. To do this, we reduce the PA gain from $g = 1$ to $g = \frac{1}{8}$ because (5-3) indicates that the loop power is proportional to g^2 . However, in doing this, (5-1) indicates that the PA efficiency reduces 8-fold, so we have to supply 8W to the PA to produce 1W in the transmitter loop. The net effect of this is that the sixth-power law for range becomes a third-power law when the efficiency of the PA is taken into account.

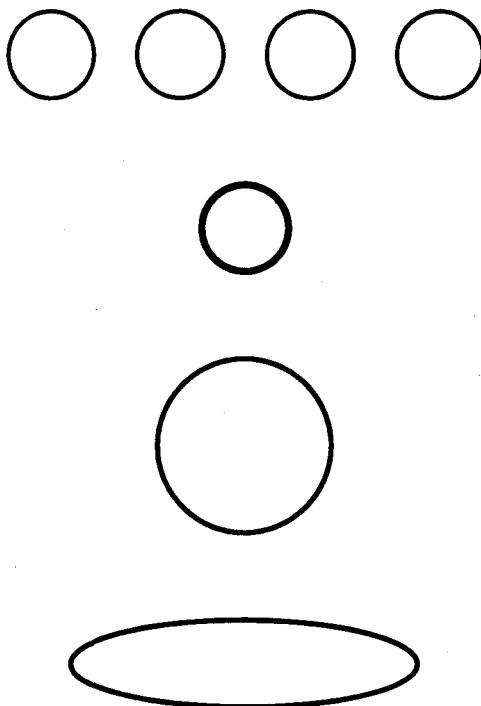
A method of eliminating this undesirable reduction of efficiency at reduced drive levels, suggested by D. Potter (*pers. comm.*) and described in (Gibson, 1992), is to control the PA rail voltage using a switched-mode power supply (SMPS). This need only have a low, sub-audio, bandwidth but will still be able to track the long term PA power demand; the design of the SMPS being much less stringent than if it were a full class-D r.f. modulator.

4.3.2.3 *Air-cored Antenna Arrays*

We can pose the question of whether there is any benefit in using multiple antenna loops, perhaps tuned to slightly different frequencies. This would help to alleviate bandwidth problems if the antennas had too high a *Q*-factor for speech, but clearly this will only work if the two antennas have very little mutual coupling. A circuit simulation using PSPICE has suggested that the coupling needs to be of the order of 0.0001 for the tuned circuits to be effectively independent. Such a coupling would require an antenna spacing of the order of twenty times the loop radius (i.e. $\sqrt[3]{10,000}$).

Figure 4-3 shows the effect of distributing the mass of the antenna system amongst several smaller antennas. These results, obtained by substitution into the equation for specific aperture (3-5), show that, given a certain mass of material, there is no benefit to be obtained by distributing it over several antennas. Even if we need to use an antenna in a confined space,

Figure 4-3d shows that it is best to construct an elongated antenna than several smaller symmetric ones. For a transmitter, the reason for this is clear – to increase the magnetic moment we must separate the ‘outward’ and ‘return’ current paths as much as possible, otherwise the magnetic field generated by the opposing current elements can cancel to larger degree. This topic was discussed in §2.2.2.3.



a) Consider, for reference, an array of N small antennas, each of radius r and mass M

b) A single antenna of the same radius and overall mass has the **same** specific aperture as the N antennas in (a) above.

c) However, a single antenna of the same overall area and mass has a specific aperture that is **larger** than the array in (a) by a factor of \sqrt{N}

d) If we take example (c) but alter the shape of the antenna to that of an ellipse, then it still has the same overall area and mass as (a) but its specific aperture is greater by a factor of $2k(k + 1)$ where k is the ratio of the radii of the ellipse. If the largest dimension of the ellipse matches that in (a) then $k = \sqrt{N}$ and, for $N \gg 1$, the advantage over (a) is approximately 2 times.

Figure 4-3 : Comparison of antenna arrays.

When we consider the mutual coupling of multiple small antennas it is possible that there could be an advantage in certain circumstances. The Q -factor or bandwidth of the array depends on the degree of coupling, and additionally, whether the array is tuned to resonance or not; but there are only limited circumstances in which we could make use of this behaviour, so we will not pursue this topic at the present time. However, we can note that the above discussion applies to air-cored antennas whereas, if the antennas are ferrite-cored, then the mutual coupling of array elements will alter the effective permeability and will therefore have an additional impact on the specific aperture. We will discuss this further in §4.4.2.

4.3.2.4 Matching the Antenna Mass

For a power drain P and a battery of mass M_e , energy storage density ρ_e , the power can be maintained for a time $\tau = M_e \cdot \rho_e / P$. Substituting this into (4-14) gives

$$D_{\max}^6 \tau \propto \Phi^2 M_e \rho_e \propto M_l M_e \quad (4-15)$$

showing that the range D_{\max} and the transmission time τ are related to the product of the battery mass and the antenna mass (written M_l above). This is a classical ‘matching’ situation where, for a given total mass of equipment, $M_l + M_e$, the product $D_{\max}^6 \tau$ is a maximum when the mass of the antenna is equal to the mass of the battery.

This result is of more use for maximising transmission time than range, because the sixth power of D_{\max} means that it is not very sensitive to changes in the ratio of the masses. And, in many systems, it will not be practical or necessary to attempt to match the mass – for example, a team of ‘sherpas’ can carry a near-unlimited mass of batteries. However in some specific systems where the total mass or volume of the equipment is critical – for example body-worn beacons or pagers, and underwater sondes – this result may have some use.

4.4 Magnetic-cored Antennas

The usual shape for an air-cored antenna is a flat disc. We could use a solenoid shape if we wanted to reduce the self-capacitance, but such a structure would, of course, be bulky. The best shape for a magnetic-cored antenna is a long thin solenoid, because this allows us to utilise the magnetic property of the material to its fullest extent, as will be discussed below. Usually we would use a ferrite as the magnetic material although soft-iron laminations might be useable at frequencies below a few hundred Hz, and a high-permeability amorphous metal might be appropriate for a very high-performance system.

Considering a solenoid with a ferrite core, this would have a very low specific aperture, due to the small diameter of the winding, were it not that the high permeability core more than compensates. It will be shown that the magnetic moment of such an antenna depends on the *core volume*, whereas the ‘efficiency’ (by which we mean the ratio of magnetic moment to power, i.e. the specific aperture) depends on the permeability and the *winding volume*. In turn, the permeability depends on the *shape factor* of the rod. A long thin rod approaches the intrinsic permeability of the material, whereas a short fat rod is *geometry-limited* in permeability.

Clearly the situation is a complex one, and it is not surprising to find that there are optimum ratios of copper to ferrite, length to diameter and so on; although these will usually have to be found by iteration, with a particular system in mind, since it is not easy to generate meaningful and succinct design equations.

We will now discuss some parameters of magnetic-cored antennas.

4.4.1 Magnetic Moment (Transmitter)

The magnetic flux density at the centre of a long thin solenoid, of length ℓ , and comprising N turns is

$$B = \mu \frac{NI}{\ell} \quad (4-16)$$

where I is the current and μ is the permeability of the material. The field is not uniform, and it is easily shown that, for a long thin solenoid, the on-axis flux density at the ends of the solenoid is half that in the centre. See, for example, (Clemmow, 1973) §4.4.2. We can also apply the Biot-Savart law to show that if the solenoid is not long and thin, but has a *shape factor* defined in terms of the length and radius as

$$m = \frac{\ell}{2a}, \quad (4-17)$$

then the field in the centre is modified by the factor $m/\sqrt{m^2 + 1}$. This variation in flux density means that the magnetic material at the ends of the solenoid is not so well utilised as that in the centre. The core shape that uses the material most efficiently is the prolate ellipsoid which, when clad with a uniform winding about its long axis, generates a uniform flux density (Burrows, 1978) §4.2. When this is so, with the flux density given by (4-16), we can write the magnetic moment (i.e. the current-area product of the equivalent air-cored single-turn) due to a thin section of the solenoid as

$$\delta m_d = \mu_r \left(\delta x \frac{N}{\ell} \right) I A(x) \quad (4-18)$$

where A is the cross-sectional area as a function of the distance along the axis, x . Substituting (4-16) we have

$$\delta m_d = \frac{B}{\mu_0} A(x) \delta x \quad (4-19)$$

which we can integrate over the length of the solenoid, x , to give

$$m_d = V_c \frac{B}{\mu_0} \quad (4-20)$$

where V_c is the volume of the core. This standard result is important, because it relates the magnetic moment of the transmitter to the ‘cost’ parameters of core volume V_c and flux density, and allows us to express the maximum m_d in terms of the saturation flux density of the core.

4.4.2 Demagnetising Factor

4.4.2.1 General Expression

In the expressions above, it was understood that μ was the intrinsic or bulk permeability of the material, but this is not necessarily the case. The *effective permeability* of the magnetic material can only approach the intrinsic permeability if the shape factor of the core is high – that is, it is long and thin. We need not analyse the effect in detail here, and will limit our observations to noting that it is a ‘fringing’ effect, as depicted in *Figure 4-4*. This effect is traditionally known as *demagnetisation*, or ‘the demagnetisation by open ends’, and is quantified in terms of a *demagnetising factor* ψ which relates μ'_r , the *effective relative permeability*, to μ_r , the (intrinsic) relative permeability of the material by the relationship

$$\frac{1}{\mu'_r} = \frac{1}{\mu_r} (1 - \psi) + \psi \quad (4-21)$$

but in most cases, the demagnetising factor is sufficiently less than unity for the approximation

$$\frac{1}{\mu'_r} \approx \frac{1}{\mu_r} + \psi, \quad \psi \ll 1 \quad (4-22)$$

to be valid. The demagnetising factor is, essentially, a figure of merit that describes the degree of under-utilisation of the magnetic material.

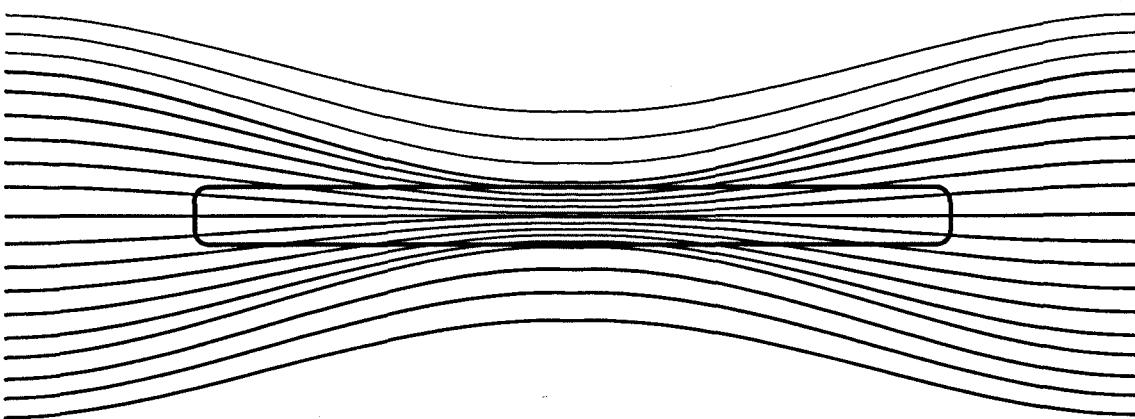


Figure 4-4 : Fringing effect, or ‘demagnetisation’ of a solenoid.

A cylindrical rod, placed in a uniform field, experiences a non-uniform flux density.

An ellipsoid would contain a uniform flux density, but the degree of concentration of the flux would depend on the shape factor.

The demagnetising factor is quoted by (Borzorth and Chapin, 1942) for a prolate ellipsoid magnetised along its long axis as

$$\psi = \frac{1}{m^2 - 1} \left(\frac{m}{2 \cdot \sqrt{m^2 - 1}} \ln \left(\frac{m + \sqrt{m^2 - 1}}{m - \sqrt{m^2 - 1}} \right) - 1 \right). \quad (4-23)$$

(Burrows, 1978) §4.2 quotes this in a slightly simpler equivalent form, and uses the reciprocal of ψ , writing

$$\frac{1}{M_d} = \frac{1}{m^2 - 1} \left(\frac{m}{\sqrt{m^2 - 1}} \ln \left(m + \sqrt{m^2 - 1} \right) - 1 \right). \quad (4-24)$$

which is conceptually useful, because M_d can then be considered as a quantity with a similar role to permeability, and the effective permeability is governed mainly by the lower of μ_r and M_d .

For $m \gg 1$, the expressions approximate to

$$\psi \approx \frac{\ln(2m) - 1}{m^2}, \quad m \gg 1 \quad (4-25)$$

which results in ψ being approximately 5% low for $m = 6$, and around 2% low for $m = 9$. We are unlikely to use antennas with a lower shape factor than this, but we can note that $m \rightarrow 1 \Rightarrow \psi \rightarrow 1/3$, giving the standard result⁶ that a homogenous sphere of high permeability, immersed in a uniform field, has an effective relative permeability of exactly 3. (For shape factors of $m < 1$ a different formula applies).

(Borzhorth and Chapin, 1942) also consider the reduction in effective permeability of a toroid containing an airgap. By performing $\int H \cdot d\ell$ around the toroid, and equating this to the source of the field, NI , we can show that there is a reduction in effective permeability given by

$$\psi \approx \frac{\alpha}{2\pi r}, \quad \alpha \ll r \quad (4-26)$$

where α is the length of the air-gap and r is the radius of the toroid. This demagnetisation is attributable to the ‘open ends’ rather than specifically to a ‘fringing’ effect and it leads us to deduce that if we introduce a small air-gap into a long thin solenoid (e.g. by constructing it from two shorter rods that are not fully butt-jointed) the resulting demagnetisation factor is of a similar form – i.e. the ratio of core length to airgap length.

4.4.2.2 Approximation for Cylindrical Solenoid

We stated, in §4.3.1, that the winding on a cylindrical ferrite rod should be concentrated over the central section. Even though *Figure 4-4* is not to scale, we can see that the central section of a rod has a higher flux density and that restricting the winding to the central section will result in a better utilisation of the copper. Just as the prolate ellipsoid has a uniform flux density when wound with a uniform winding, we could surmise that a cylinder may have a uniform flux density when the winding has a cross-section of an ellipsoid or similar shape. Restricting a uniform winding to a central section of the solenoid is only a crude approximation to that of a uniformly clad prolate ellipsoid and we will not attempt to justify it analytically.

6 See for example (Ramo, Whinnery and Van Duzer, 1984) §7.18.

However we will note that the validity of any assumptions we may make about inductance, flux density and so on do not prevent us from making comparisons *between* antennas.

4.4.2.3 Ferrite Tubes

An argument presented by (van Suchtelen, 1952) shows that, for the case of a tube of magnetic material, the demagnetisation factor is still governed by the outside dimensions. The flux density still obeys the relationship $B = \mu' H$, but the flux, being $\phi = A B$, depends on the cross-sectional area actually occupied by the ferrite core (the assumption being that $\mu' \gg \mu_0$). We can make appropriate modifications to the expressions for transmitter and receiver expressions to allow for this. There would appear to be no particular reason why we would choose to use a tubular material for its performance, although it may be easier to handle, should we wish to join tubes together. Tubes are often commercially available and may, in some circumstance, be easier to obtain than a rod.

4.4.2.4 Optimum Shape Factor

Consider a long thin solenoid. Its ‘inherent’ specific aperture (by which we mean without a magnetic core) will be very low, but the magnetic core makes up for this. Suppose we decrease the shape factor whilst keeping the ratio of winding parameters to core parameters (as shown in *Figure 4-2* constant. The inherent specific aperture will increase, but the effective permeability will decrease, and there will be an optimum shape factor at roughly the point where the effective permeability changes from being material-limited to geometry-limited.

Alternatively, suppose that some of the mass of the winding is exchanged for core mass, whilst keeping the length constant. The shape factor of the core will reduce, so its permeability will reduce. On the other hand its diameter will increase, so the inherent specific aperture of the winding may increase, although it is also likely to decrease because we have removed some of its mass. Clearly the situation is a complex one, but preliminary investigation with a spreadsheet has shown that there may be an optimum ratio of winding mass to core mass in some circumstances, but these are difficult to express succinctly. In many systems, the design will be governed by the material available and so this discussion may be academic in most situations.

4.4.3 Specific Aperture and Inductance

The expression for specific aperture given in (4-10) assumed that the antenna had a uniform cross-section, as well as a uniform flux density. To derive the specific aperture for a prolate ellipsoid requires us to integrate over the length of the ellipsoid, and we must perform a similar exercise to obtain the winding resistance and inductance. These expressions are not required

unless we decide to utilise a uniform winding on a prolate surface. For practical work we will almost certainly be using a cylindrical core.

4.4.4 Power Dissipation (Transmitter)

4.4.4.1 Winding and Core Losses

The winding will suffer from losses due to the skin and proximity effects (see §4.7) and from self-capacitance. However, a ferrite rod may well experience far lower winding losses than an air-cored loop because, not only is the winding smaller in volume, but it may well be in a single layer rather than a large bundle with an inherently high self-capacitance.

However, ferrites and other magnetic materials are subject to power dissipation caused by hysteresis losses and eddy current losses within the magnetic material. These ‘core losses’ can be high – for a typical ferrite, such as material grade 3C85, they amount to around 180mW/cm³ (at a flux density of 0.2T and a frequency of 25kHz) so, for the 90cm³ array discussed above, they amount to 16W. This is the strongest argument for not using ferrite rods in the transmitter antenna, other than at very low frequencies but, as we will discuss briefly in an example in §4.8.3, amorphous metals cores can have losses of less than 10% of a typical ferrite, although they are extremely expensive.

4.4.4.2 Linearity

Ferrite or iron powder cores may saturate or, at least, be non-linear. Given the stringent limits of spurious emissions, this needs to be analysed carefully. The distortion can be reduced by feedback, of course, but it can also be reduced in an open-loop fashion. Consider a material where the permeability μ is a function of the applied field strength H . We can write the flux density in the solenoid as

$$B = \mu H = \frac{\mu NI}{\ell} \quad (4-27)$$

where ℓ is the length and NI is the energising amp-turns. The inductance is defined as $L = N\phi/I$ where ϕ is the flux, and so the magnitude of the inductive voltage U_x can be written

$$U_x = L \frac{di}{dt} = |j\omega N \phi| = \omega NAB \quad (4-28)$$

which leads us to write the magnetic moment of the antenna as

$$m_d = \frac{U_x \ell}{\omega \mu_0 N} \quad (4-29)$$

which shows that if we drive the solenoid from a voltage source instead of a current source we will have the required linear relationship between the magnetic moment and the driving signal⁷.

4.4.5 Antenna Arrays

As noted earlier, an array of ferrite rods that are not magnetically isolated will suffer from a reduced effective permeability and so it is necessary to examine a number of configurations, as shown in *Figure 4-5*, below. These configurations of N rods can be described as follows.

- a) ***Co-axial:*** The rods are joined co-axially into one long rod and the winding placed on the central section. Since $\Phi \propto r\sqrt{M}$ and the mass has increased by N times then Φ increases by \sqrt{N} times; and since $B \propto \ell / M$, with the length and mass both increasing by N times, then the bandwidth remains unaltered – apart from any change caused by a change in the effective permeability.
- b) ***Parallel:*** The rods are formed into a single short fat rod. In practice we would use either a rod of these new dimensions or we would bundle the smaller rods together before placing the winding over the assembly. Here, the radius increases by \sqrt{N} times, leading to an overall increase in Φ of N times. Since the length is unaltered, the bandwidth decreases by N times.
- c) ***Isolated:*** We construct N antennas using a single rod each, and placed them far enough apart that they do not interact. For the antennas to be isolated there must be no mutual inductance and no additional demagnetisation factor. It seems reasonable to assume that this can be achieved if the rods are spaced by a distance equal to their length⁸.
- d) ***Semi-isolated (co-planar):*** We take the previous example but assume that the rods are close enough for full coupling. In this situation, if the windings are connected in series then the inductance will increase by N^2 times, leading to a decrease in bandwidth of \sqrt{N} times. A similar result occurs for parallel connection of the windings.
- e) ***Semi-isolated (coaxial):*** Alternatively, if the rods were in a co-axial alignment, then a small air-gap between the rods would suffice to reduce the effective permeability to that of an individual rod, whilst maintaining nearly full inductive coupling⁹.

Additionally, the shape factor plays a large part in these comparisons because it can have a significant effect on the effective permeability. It is tabulated, for the above configurations in

7 The relationship between U_X and the driving voltage depends, of course, on the Q -factor and the tuning.

8 For example, if two rods, each 200mm \times 8mm diameter, were spaced by 50 times their radius (i.e. 200mm) we could expect the coupling to be of the order of $1/50^3$ of the field at the centre of the rods.

9 For example, the demagnetising factor of a single 200mm \times 8mm diameter rod is around 1/200. A high-permeability rod will have μ' shape-factor-limited to 200. A separation of 1mm between several coaxial rods will reduce the otherwise much larger μ' to a figure of the order of 200.

Table 4-3 below, which also shows the bandwidth and specific aperture for the configurations. The table covers only a general case; it is not possible to state the exact change in μ' , because this depends on the value of the intrinsic permeability of the material. However, we can make an estimate of performance by approximating the demagnetising factor as $\psi \approx 1/m^2$ (see equation 4-25). The specific aperture Φ must be multiplied by μ' to give the magnetic specific aperture (MSA) and, although the table shows an increase in Φ over that of a single rod, this can be more than offset by the reduction in μ' caused by the change in shape factor. We can see that semi-isolated co-planar rods (*Figure 4-5d*) provide a particularly unfavourable alignment. We will make further comment when we consider a specific example in §4.5.

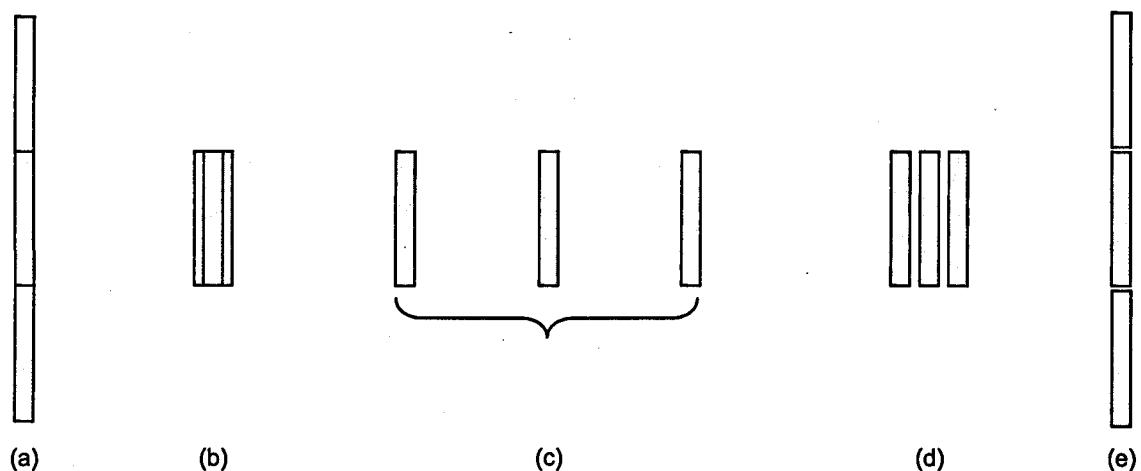


Figure 4-5 : Alignments of multiple ferrite rods.
(a) Co-axial, (b) parallel, (c) isolated, (d) & (e) semi-isolated.

	Single Rod	Coaxial	Parallel	Isolated	Semi-Isolated (coplanar)	Semi-Isolated (coaxial)
Increase in shape factor	1	N	$1/\sqrt{N}$	1	$1/\sqrt{N}$	1
Increase in specific aperture	1	\sqrt{N}	N	\sqrt{N}	\sqrt{N}	\sqrt{N}
Increase in demagnetising factor (1)	1	$\approx 1/N^2$	$\approx N$	1	$\approx N$	1
Decrease in bandwidth (2)	1	1	N	1	N	1

Table 4-3 : Change in specific aperture for five configurations of N rods.
A single rod of shape factor m is shown for comparison. Note that the specific aperture must be multiplied by the effective relative permeability, which is a function of the shape factor.

- (1) Increase in demagnetising factor is approximate – see text.
- (2) Decrease in bandwidth refers to the contribution due to length & radius. There will be an additional contribution from the change in demagnetising factor.

4.5 Air-core v. Magnetic-cored Antennas

We have seen that the inherent very low specific aperture of an air-cored solenoid is increased by the insertion of a magnetic core. The important question is how this performance compares with that of a larger air-cored antenna of equivalent mass or power consumption. Unfortunately it is not easy to make generalised statements, not least because the choice depends on the relative importance of several factors, including

- i) Mass
- ii) Power (for transmitter)
- iii) Size (radius or length)
- iv) Bandwidth

and the requirements of a speech communications system may be different to those of a beacon or radiolocation device. To demonstrate the point, we will make some very brief comparisons with the aid of a spreadsheet.

4.5.1.1 *Air-cored Loop*

We will consider four antennas that would be suitable for cave communications – a small heavy loop 450mm in diameter, similar to that used for radiolocation work at under 1kHz; a 1m diameter loop on a collapsible frame, similar to that used for speech communication at 87kHz, and two larger installations with 2m and 10m diameter loops. Such large loops are not usually used for cave communications, but mining installations have used loops hundreds of metres in diameter. See *Table 4-4*, below.

The immediate observations are that the smaller radiolocation loop has to have a high mass in order to have the same performance as the 1m speech loop, and that it has a much lower bandwidth. The 2m speech loop has double the specific aperture of the 1m loop, for the same mass, and therefore, when used as a transmitter it will increase the range by 25% (i.e. range increases by $\sqrt[3]{2}$). When used as a receiver, the improvement in range will depend on whether the antenna is large enough to swamp the receiver noise. If it is then there will be no improvement, but if the antenna is ‘small’ then there will be a 25% increase in range. (See §3.4.1.1).

4.5.1.2 *Ferrite-cored Solenoid*

We will consider a small rod of diameter 8mm and length 200mm, with a copper winding as described in §4.3.1. Such rods are readily available in a MnZn ferrite, type 3C85, with an initial permeability of around $\mu_i = 2000$. On its own, this is quite a small antenna, but we will look at arrays of multiple such rods, as described in §4.4.5, above and *Table 4-5*, below.

Parameter	Loop 1	Loop 2	Loop 3	Loop 4
Loop Diameter, m	0.45	1	2	10
Wire Diameter, mm	0.315	0.2	0.2	0.2
Strands	1	7	7	7
Number of Turns	1016	34	17	4
Winding mass, g	1000	210	210	247
Specific Aperture, $\text{m}^2/\sqrt{\Omega}$	9.1	9.2	18.5	100.1
Bandwidth, Hz	20	185	346	1280
Power, W	10	10	10	10
Moment, Am²	29	29	58.4	317

Table 4-4 : Comparison of typical air-core loops.
Parameters in bold show values derived from the data shown in un-emboldened type.

One rod	Three rods					Nine Rods					
	a	b	c	d	e	a	b	c	d	e	
Solenoid Diameter, mm	8	8	14	—	—	8	8	24	—	—	
Solenoid Length, mm	200	600	200	200	200	600	1800	200	200	200	
Solenoid mass, g	48										
Shape Factor	25	75	14	25	14	25	225	8	25	8	
Winding/solenoid length	1/3										
Winding thickness/sol. dia	1/8										
Winding/core ratio	0.31										
Effective winding radius, mm	3.56	6.16					10.67				
Winding mass, g	15										
Total Mass, g	63	190	569	
Intrinsic Permeability	2000										
Effective Rel Permeability	194	824	84	194	84	194	1664	38	194	38	
Specific Aperture, $\text{m}^2/\sqrt{\Omega}$	0.018	0.030	0.053	0.030	0.030	0.030	0.053	0.158	0.053	0.053	
Magnetic Specific Aperture	3.4	25.0	4.4	5.9	2.6	5.9	87.5	5.9	10.2	2.0	
Bandwidth, Hz	86	20	65	86	65	86	10	49	86	49	
Saturation Flux density, mT	400										
Total Core Volume, cm³	10	30	90	
Max. Magnetic Moment, Am²	3.2	9.6	9.6	9.6	9.6	9.6	28.8	28.8	28.8	28.8	
Maximum Power, W	0.89	0.15	4.67	2.66	14.02	2.66	0.11	23.6	7.99	213	
										7.99	

Table 4-5 : Comparison of different alignments of ferrite-cored solenoid antennas.
Data is shown for a single rod, and for cases (a) to (e) for arrays of three and nine rods.
Parameters in bold show values derived from the data shown in un-emboldened type.

Table 4-5 makes use of the data presented in Table 4-4 and the analysis of §4.3.1. It was derived from a spreadsheet, which is included on the bound-in CD-ROM (See appendix A9). We can see that the MSA of a single rod is enhanced many times by the co-axial alignment of rods (*case a*). Of the remaining cases, the isolated rods (*case c*) and semi-isolated coaxial (*case e*) have a similar performance but we can observe that if the same material is used to make a

single short, fat rod (*case b*) then the performance is worse than that of the isolated rods (*case c*). However, if we use isolated rods we must be careful not to approach the configuration of a 'short, fat rod' because the performance (*case d*) reduces to *below* that of a single rod. If we cannot avoid close proximity of the rods, it will be best to keep them semi-isolated in a co-axial arrangement (*case e*).

Use as Receiver: As a receiver, a coaxial array of three such rods out-performs the 450mm loop of *Table 4-4*, in that it has a higher specific aperture for a lower mass; and a bandwidth useable for radiolocation. However, to make useful comparison with the 1m and 2m loops we need to specify the purpose of the antenna. If it were used in a tuned receiver, the solenoid would have too low a bandwidth for speech. We will discuss the operation of an untuned receiver antenna in chapter 6. Using the antenna untuned, it would out-perform both the larger loops.

Use as Transmitter: The MSA tells us the performance as a receiver but, for a transmitter, the magnetic moment is limited by the saturation flux density of the material. For power efficiency the co-axial alignment is clearly the best but, with three rods, the magnetic moment is limited to under 10Am^2 . We can make small changes to the MSA by increasing the winding volume¹⁰ and, for a transmitter, using the appropriate value for the amplitude permeability¹¹ but if a substantially higher magnetic moment is required then it becomes necessary to increase the core volume or use a material with a higher saturation flux density. However, to do this and achieve a usable bandwidth may not be possible.

We can see that to achieve the same magnetic moment as the 1m loop achieves for a power input of 10W we need a core volume corresponding to nine of the specified rods. Such an antenna requires only a very low power for the co-axial array, but it has a larger mass and a lower bandwidth than the air-cored loop. Whether this makes a more suitable antenna than the air-core loop depends on the intended purpose. However, the low bandwidth (i.e. high *Q*-factor) is almost certainly not achievable due to core losses, as discussed below.

In conclusion, we can see that the shape factor of a ferrite cored solenoid, and the diameter of an air-cored loop are both highly significant, and that general comparisons are therefore difficult. We can observe that a powerful transmitter will benefit from a large diameter loop or a very long solenoid but that, in a portable system, there is always an inherent limitation in the magnetic moment we can achieve from a solenoid of reasonable dimensions.

10 The figure of $t / 2r = 1/8$ (equation 4-12) is low. (Burrows, 1978) uses 0.4 in one example. Also, we have used $\ell_w / \ell_c = 1/3$ whereas Burrows uses ≈ 0.45 in one example [fig. 57].

11 For 3C85 ferrite, the initial relative permeability – i.e. dB/dH at $H = 0$ – is around 2000. The amplitude relative permeability – B/H at saturation flux density – is around 5000.

4.6 Coupling to the Ground

The equations presented in chapter 2 described the propagation of the fields, but they did not describe the process of coupling the signal to the ground. In a conventional radio we can describe this in terms of the radiation resistance of the antenna. In our, essentially, induction radio system it is described by the *external resistance* which was defined in §3.4.3, which plays a similar role to radiation resistance.

However, external resistance does not, in itself, allow us to describe the effect on the apparent inductance of a loop that is coupled to a conducting medium. Using conventional lumped circuit theory, we can describe the apparent impedance of the loop in terms of the flux that is coupled back from the ‘secondary’ (i.e. the rock) to the ‘primary’ (i.e. the antenna) but we then need to perform a detailed analysis to calculate this field. The equations presented in chapter 2 are insufficient because that do not cover the case where the distance to the field point is zero. The problem is analysed in a different way by (Kraichman, 1962), but we have not pursued this line of enquiry because practical experiments by J. Hey¹² have shown that the effect appears to be small. However, in a high-*Q* tuned antenna we would need to address this potential de-tuning problem.

At this point we may ask whether it is possible to achieve a directive transmission into the earth, and whether we can save power by contriving to restrict the magnetic field to a directed beam. This, of course, is not possible with the mainly induction fields we are using; and since the fields do not propagate energy into free space, it does not matter that we generate a large magnetic field in the air. However, it is possible to achieve a directive transmission if the loop antenna is of a resonant size (Smith and An, 1983), but such antennas could not form part of a portable communications system

4.7 Proximity and Skin Effects

The *skin effect*, whereby an electric field decays exponentially as it penetrates a conducting medium is well known. The skin depth formula derived in chapter 2 applies to this situation, and it can be shown that, provided a wire is several skin depths in diameter¹³, the current flow is *as if* all the current were conveyed, uniformly, in a surface layer one skin depth deep. For example, the skin depth in copper at 100kHz is 0.2mm so that a wire that was, say, 2mm in diameter would have a resistance at 100kHz that was approximately 2.5 times its d.c. resistance. Clearly this has an implication for how we construct the antenna and, for a given

12 Carried out at the request of the author.

13 If this approximation cannot be made then a more complete analysis is required. See, for example, (Ramo *et al.*, 1984) §4.5 ‘Impedance of round wires’.

mass of winding, many turns of thin wire may be better than a few turns of thicker wire. However, this does not take into account the effect of self-capacitance, or the proximity effect, now to be described.

The skin effect can be explained by considering that the current causes a magnetic field inside the wire that gives rise to eddy currents, which oppose the flow of current in the centre of the wire, and re-enforce it on the outside. The *proximity effect* is related to the skin effect, but is caused by an *external* field. In a pair of parallel wires carrying current in the same direction, the proximity effect will cause the lines of current flux in the two wires to move apart. In a bundle of wires this has the effect of forcing the current to the outermost edges of the bundle.

Although similar to the skin effect, the proximity effect operates independently of it. If the wires in the bundle are very close together then the current distribution will be like that of a single larger wire and there is no specific effect due to proximity. This is also the case if the wires are well-spaced. However, for situations between these two extremes, it was shown by (Butterworth, 1926a; b; c; d) that there was an additional effect. Butterworth explained that the skin effect was reduced in thin wires, but that the proximity effect was reduced in thick wires and that there was an optimum spacing. His results were summarised by (Austin, 1934) and by (Smith, 1940). Butterworth's results were discussed by (Watt, 1967) but Canadian cave radio experimenter Ian Drummond believes that Watt misinterpreted some of Butterworth's results. Based on a study of Butterworth's formulas, Drummond commented (Drummond, 1992),

"It is certainly worth noting that the a.c. resistance of a [loop] antenna can be as much as ten times the d.c. resistance of the same piece of wire, and that most of the increase comes from the effect of the antenna field on the wire, not from the effect of the isolated wire's field on itself (which is what most people think of as the 'skin effect', I think). So any antenna analysis must take these effects into account to be even half (or tenth)-way accurate."

The proximity effect is also referred to briefly by (Balanis, 1997) §5.2.3 and was analysed by (Smith, 1972a; b) who found it to be a significant factor in the performance of electrically small multi-turn loop antennas. There is no doubt that this important effect clearly needs further study, as it will be beneficial if we can optimise the antenna winding in such a way that we can reduce the proximity effect. However, it must be noted that none of these analyses of the proximity effect consider losses due to self-capacitance. It is probable that this will also give rise to a significant effect although Drummond's comment, above, was verified for an antenna with a low self-capacitance loss.

4.8 Exotic Antennas

We have described some conventional types of loop antennas, but there are some novel or 'exotic' schemes that should also be given consideration. These include arrays of ferrite rods,

specialised magnetic materials, permanent magnets as transmitters, magnetometers as receivers and toroidal antennas.

4.8.1 Toroidal Antennas (Transmitter)

There has been some interest, in recent years, in the idea of using a toroid – that is, a solenoid bent into a doughnut shape – as a transmitter. We had considered this as a means of favourably altering the inductance to resistance ratio of a transmitter loop, but could draw no definite or useful conclusions (Gibson, 1994). The toroid as a non-radiating structure for exciting a conducting half-plane was considered by (Wait, 1995a; b). It has been considered as a radiating structure by several authors, (Beers, 1996; Carron, 1995; Jennison, 1995), for which the term *anapole antenna* has been used. Prof. Jennison has also described a practical antenna for the 14m band (Jennison, 1994a; b), which was evaluated by (Dodd, 1994). A more complex variation – the contrawound toroidal helical antenna – was developed at CIRA¹⁴, described by Smith¹⁵, and was the subject of a patent (van Voorhies and Smith, 1992).

We could analyse a toroid from first principles simply by integrating the magnetic vector potential, as described in appendix A2.2 and we could, perhaps, make the process easier by approximating the toroid as a number of series connected loops. Wait, however, takes an approach based on the concept of a ‘magnetic current’. Just as an electric current (defined as rate of change of electric charge dq/dt) flowing in a loop generates a magnetic field, so a magnetic current (defined as rate of change of magnetic flux $d\phi/dt$ and with the dimension of volts) flowing in a loop generates an electric field. Wait showed that the horizontal anapole therefore behaves like a vertical electric dipole. One condition for this is that the anapole can be treated as a point source. A second condition, which Wait did not explicitly state, is that the magnetic current is much larger than the electric current. If this were not the case then a significant vertical magnetic dipole would be superimposed on the field.

The horizontal anapole therefore provides an efficient method of synthesising a vertical electric dipole, which we could otherwise only achieve with a tall monopole¹⁶. Wait observes that there are advantages to using a vertical electric dipole for geophysical measurements because the magnetic field is entirely transverse, and will interact strongly with flat-lying, thin resistive layers.

¹⁴ Centre for Industrial Research Applications, College of Engineering and Mineral Resources, West Virginia University, USA.

¹⁵ J.E. Smith, *Antennas*, web page at www.cira.wvu.edu/cur_proj.html#Antenna (no longer available but the information is traceable via [google.com](https://www.google.com)).

¹⁶ A brief analysis of a monopole supported by a balloon was given by (Gibson, 1997a). Using the concept of ‘specific length’ in a similar fashion to ‘specific aperture’, it was shown that, for the same lifting power, a short thick wire (or bundle) was preferable to a long thin wire.

From a communications point of view, the transverse field will be captured by a vertical loop antenna edge-on to the axis of the transmitter. This has several advantages:

- i) The receiver is always easy to orientate, unlike the situation for a VMD transmitter where the receiver has to lie at an angle when it is off-axis.
- ii) The simple orientation follows the scheme used for conventional (far-field) radiolocation, where the receiver detects the magnetic far-field component of the signal from a VED. Radiolocation is therefore simpler.
- iii) The anapole has the same field pattern as an electric dipole, so the magnetic field strength falls off as the inverse square of distance, rather than the inverse cube.

Wait defines magnetic current, K , as the rate of change of flux in the toroid, which we can write as

$$K = \frac{d\phi}{dt} = j\omega\mu N_c \pi a^2 \quad (4-30)$$

where I is the current, N_c is the number of turns per unit length on the toroid, and a is the radius of the flux tube. Wait equates this to the electric dipole moment by

$$I\ell = j\omega\epsilon K\pi r^2 \quad (4-31)$$

where r is the radius of the toroid. From these expressions we can derive the specific length of the antenna (i.e. the electric dipole moment, relative to \sqrt{R} , where R is the resistance of the winding) and the specific aperture, treating the toroid as a single turn loop antenna. The expressions are

$$\Phi_m = \frac{1}{2\pi a N_c} \cdot \frac{1}{2} r \sqrt{M \frac{\sigma}{\rho}} \quad (4-32)$$

which demonstrates, as expected, that a practical toroid (high N_c) has a very low specific aperture, and

$$\Phi_e = \mu_r k_0^2 \frac{1}{2} a \cdot \frac{1}{2} r \sqrt{M \frac{\sigma}{\rho}} \quad (4-33)$$

where k_0 is the wave number in free space and μ_r is the relative permeability of the core material. The presence of the k_0^2 term means that, at low frequencies, the anapole antenna does not generate a large magnetic field. At higher frequencies it would appear to be an attractive device, but further analysis is required. We can observe that a toroid would enable us to preserve the high intrinsic permeability of a magnetic material but that, as with the ferrite rod antennas discussed earlier, the dipole moment would be limited by the saturation flux density.

4.8.2 Rotating Permanent Magnet (Transmitter)

A permanent magnet, slowly rotating about an axis perpendicular to its dipole moment will produce a magnetic field that is the sum of two perpendicular current loops driven in phase quadrature. With simple equipment we might be limited to a rotational speed of 12,000 rpm which would allow us to generate a magnetic dipole field at 200Hz. Clearly this is only suitable for a radiolocation beacon, and not for data transmission or speech communications. There is, however, a possible advantage to this scheme over that of a current loop or solenoid.

4.8.2.1 Magnetic Moment

The magnetic moment of a permanent magnet is described by a similar formula to that which governs the moment of a solenoid, (4-20), and we can state that

$$m_d = V_c \frac{B_r}{\mu_0} \quad (4-34)$$

where B_r is the **remanence** of the magnetic material. Although strictly, this formula only applies when all regions of the magnetic material experience the same flux density and the material has a good shape factor.

A high-performance neodymium-iron-boron (NdFeB) magnet has, typically, a remanence of 1.22T. This is several times larger than the saturation flux density of a ferrite material, which is typically 0.4T. (For comparison, soft iron has $B_{sat} = 2.2T$). In addition, there will be no losses due to hysteresis, eddy currents, self-capacitance or PA efficiency, and no problems in tuning the antenna to resonance. The only power we need to put into the system is that needed to spin the material, using a small motor. A small slab of NdFeB, measuring 50mm × 25mm × 12.5mm, and costing around £70, would present a magnetic moment of 15Am² which is similar to the moments in the examples of §5.4.2.

4.8.2.2 Demagnetising Factor

A permanent magnetic material is characterised by its remanence – i.e. the flux density for zero applied field strength – but this flux density can only be achieved in a magnet with a low demagnetising factor, ψ . Just as with a non-permanent magnetic material, we can derive an effective permeability that describes the ratio of B to H within the material. This ‘operating point’ must lie on the B - H hysteresis curve for the magnet, and the effect of a high demagnetising factor may be to force the operating point to be close to the point of coercivity (that is, the applied H field for which B reduces to zero). This makes the magnet susceptible to external fields and means that its magnetic moment may not be stable. This is illustrated in **Figure 4-6** below.

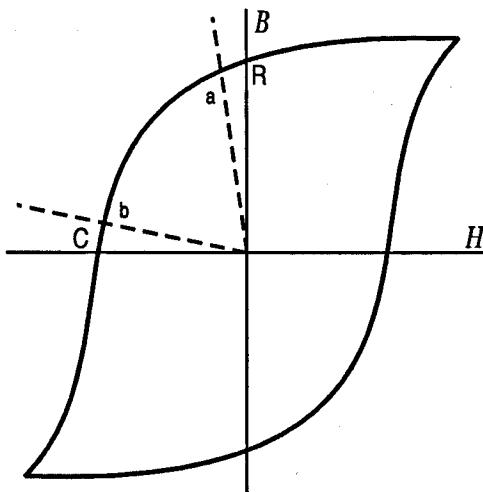


Figure 4-6 : B-H hysteresis curve for a permanent magnet.

The remanence of the material is shown at point R; the coercivity at point C.

Load line (a) shows the value of magnetic flux that is sustainable at a demagnetising factor of 0.15 (i.e. $H/B = 0.15$). Load line (b) shows that at a very high demagnetising factor – 4.7 in this case – the available flux density is low.

High-strength NdFeB magnets are available for use in high-performance servo motors and are usually supplied as slabs with a very low shape factor. This does not matter in a motor application because the magnets are part of a larger magnetic circuit, with only a small airgap. In our application we would need to ensure that the overall shape factor was high – either by using several slabs in a co-axial alignment, or by incorporating a ‘keeper’ to increase the shape factor.

4.8.2.3 Screening

The flux density of the earth’s magnetic field is around $50\mu\text{T}$ corresponding to a field strength of 40A/m . This is the field strength we would have at a distance of 400mm from the 16Am^2 source in the example above¹⁷. If screening was a problem during storage and transport of the device, we could observe that a suitable multi-dipole arrangement of 2^N magnets would produce a moment that was 2^N times as great, but which reduced as the $(N + 3)$ th power of distance. Spinning this appropriate configuration would result in a signal at 2^N times the frequency, but which probably reduced too rapidly with distance to be of use.

4.8.2.4 Radiation and External Resistance

It is interesting to consider how the effects of radiation and external resistance manifest themselves with a rotating magnet. The following discussion is somewhat heuristic, but we believe it to capture the essence of the problem. For a conventional magnetic dipole we described, in chapter 2, how the field could collapse and reform every half-cycle and that, if we

¹⁷ The UK parcels distribution service *Parcelforce* specifies a limit on the field strength of materials they will convey. Curiously, this appears to be given [*Parcelforce Guide to Services*, April 2002] as 0.159A/m at a distance of 2.1m from a parcel, which is only a small fraction of the earth’s field.

consider the field lines to be restricted to moving at the speed of light, there was a limiting distance, at $\lambda/4$, where the field could not collapse in time, and so a radiation field would be generated. The analogy for a rotating magnet would be that the field lines could not move through space faster than c and so we could expect a transition zone to occur at $\lambda/2\pi$ where the radial velocity would be c .

The manifestation of this radiation field would be a force tending to oppose the rotation of the magnet. The external resistance caused by eddy currents in the rock could be expected to result in a similar retarding force, whilst the ‘true’ frictional force of the bearing would be the analogue of the ohmic resistance of a wire loop, and would provide a figure of merit against which to compare the other effects. Entertaining though this speculation is, the effects are, of course, so small as to be immeasurable.

4.8.3 Specialised Magnetic Materials

Ferrite rods are ubiquitous, and there are several classes of ferrite ideally suited to the range of frequencies we will be using. However, a range of materials known as *amorphous metals* has been developed for use in switched-mode power supplies and similar power devices, which has superior properties to ferrites. Amorphous metal tape is wound into a range of C, E and I core shapes. None are ideally suitable for our solenoidal applications, but some customisation would be possible in a commercial application. The material itself is fairly expensive, but it does have a high permeability and a high saturation flux density.

AlliedSignal (now part of Honeywell) manufactures a range of cores using *Metglas*¹⁸ amorphous alloys, which have an extremely high permeability, high saturation flux density and very low core loss. Some characteristics are summarised in the table below, which includes a standard ferrite for comparison.

From (4-34) we can see that with a saturation flux density of 0.5T we would need a core volume of 75cm³ to achieve a magnetic moment of 30Am². Considering the 2714AF material, the core loss¹⁹ of 1.2W at 25kHz might be acceptable if the winding loss were of a similar magnitude. This level of winding loss could be achieved with a rod of length 370mm and diameter 16mm²⁰. Of course, such a rod would have a permeability that was geometry-limited, and we would not be making use of the high permeability of 90,000. In addition, the low bandwidth (or high *Q*-factor) would make this antenna suitable only for very low frequency work.

18 www.metglas.com

19 We use the rough figure from *Table 4-6*, but a formula is available from the product datasheet.

20 Using $t/2r = \frac{1}{2}$, $\ell_w/\ell_c = \frac{1}{3}$

With a saturation flux density of 1.5T we would need a core volume of only 25cm³ to achieve a magnetic moment of 30Am². Considering the SA1 material, the higher core loss of 9W would mean that this material was suitable only for very low frequency work. This particular material has only a low relative permeability and so it is possible to demonstrate the occurrence of an optimum shape factor. In this particular example, the MSA is at a maximum for a core diameter of around 8mm, for which the winding dissipation is around 13W. If the diameter is increased the antenna becomes geometry-limited, but if the diameter is decreased the effective permeability cannot increase high enough to offset the reduction in diameter. The SA1 material allows us to utilise a core volume of only a third of the 2714AF material, but it requires 10 times as much power for the winding, as well as having higher core losses.

Properties	2714AF cobalt-based amorphous metal	SA1 iron-based amorphous metal	3C85 MnZn ferrite
Ribbon thickness	18 µm	22 µm	n/a
Density	7590 kg/m ³	7180 kg/m ³	4800 kg/m ³
Saturation Flux Density	0.55 T	1.56T	0.40 T
Permeability (typ)	90,000 ±20%	250–285	2000
Core loss at 0.2T, 25kHz	16 mW/cm ³	370 mW/cm ³	180 mW/cm ³
Availability	Small toroids to 35 x 10mm	Toroids, C-cores to 106 x 85 x 85mm	widespread

Table 4-6 : Comparison of magnetic materials.

4.9 Concluding Remarks

We have considered the design of induction loop antenna systems and we have made a number of comparisons between the use of air-cored loops and ferrite-cored solenoids. It is apparent that the transmitter and receiver may benefit from different antenna designs; with the transmitter being a large air-cored loop and the receiver a smaller, long and thin solenoid. The reason for this distinction being that the magnetic moment of a ferrite-cored transmitter is limited by the saturation flux density of the material and the hysteresis losses in the core.

We discussed skin and proximity effects and noted that further investigation was needed in this area. The winding losses in an air-cored loop are likely to be larger than those in a ferrite-cored solenoid, making an overall comparison of the losses necessary in a full design.

We demonstrated that an antenna made from an array of ferrite rods had a performance that was highly dependent on the physical arrangement of the rods. We can see that it would be possible to make a very sensitive receiver antenna if it were long and thin but, in a practical system we would need to pay attention to the assembly of the antenna, and address issues

concerned with portability. These issues are also important for an air-cored loop, but it is a simple matter to assemble a cable onto a collapsible framework.

Unfortunately, we cannot make any clear-cut statements about the design of the antenna because it depends to a large extent on the requirements of the communications system, which we have not specified at this stage.

The results of this chapter will be used in the further discussion of transmitter and receiver antennas in the next two chapters.

4.10 References²¹

- Austin, B. B. (1934). The Effective Resistance of Inductance Coils at Radio Frequency. *Experimental Wireless & Wireless Engineer*, January 1934, 12-16.
- Balanis, C. A. (1997). *Antenna Theory: Analysis and Design*. New York: John Wiley & Son.
- Beers, Y. (1996). Anapoles - a Different View on Radiation from Toroidal Inductors. *Communications Quarterly*, Summer 1996, 67-71.
- Belrose, J. S. (1992). *VLF/LF Transmitting Antennas*. Conference Proceedings 529 on ELF/VLF/LF Radio Propagation and Systems Aspects: Advisory Group for Aerospace Research and Development; NATO.
- Borzorth, R. M. and D. M. Chapin (1942). Demagnetisation Factors of Rods. *Journal of Applied Physics* 13, 320-326.
- Burrows, M. L. (1978). *ELF Communications Antennas*. [IEE Electromagnetic Waves Series: 5]. Stevenage: Peter Peregrinus.
- Butterworth, S. (1926a). Effective Resistance of Inductance Coils at Radio Frequency - Part 1. *Experimental Wireless & Wireless Engineer*, April 1926, 203-210.
- (1926b). Effective Resistance of Inductance Coils at Radio Frequency - Part 2. *Experimental Wireless & Wireless Engineer*, May 1926, 309-316.
- (1926c). Effective Resistance of Inductance Coils at Radio Frequency - Part 3. *Experimental Wireless & Wireless Engineer*, July 1926, 417-424.
- (1926d). Effective Resistance of Inductance Coils at Radio Frequency - Part 4. *Experimental Wireless & Wireless Engineer*, August 1926, 483-492.
- Carron, N. J. (1995). On the Fields of a Torus and the Role of the Vector Potential. *American Journal of Physics* 63(8), 717-729.
- Clemmow, P. C. (1973). *An Introduction to Electromagnetic Theory*. Cambridge: Cambridge University Press.
- Dodd, P. (1994). Evaluation of the G2AJV Toroidal Antenna. *Radio Communication*, August 1994, 33-37.
- Drummond, I. (1992). Letters: Antenna Design. *CREGJ* 10, 5.
- Gibson, D. (1992). Cave Radio PA Design and a Digital PWM DSB Modulator. *CREGJ* 10, 16-24.

21 References to *CREGJ* are to the *Cave Radio & Electronics Group Journal*; a non-peer-reviewed low-circulation magazine published by a special interest group of the *British Cave Research Association*, bcra.org.uk. The CREG journal is filed at the British Library as ISSN 1361-4800.

- (1994). A Toroidal Loop Antenna. *CREGJ* 16, 3-4.
- (1997a). Cave Radio Notebook 15: 73 kHz Monopoles. *CREGJ* 28, 22.
- (1997b). Ground Shorts with a Single-Wire Telephone. *CREGJ* 28, 10-13.
- (1997c). High-Z SWTs. *CREGJ* 27, 6.
- (1997d). The Resistance of Ground-Electrode Arrays. *CREGJ* 29, 26-27.
- (1997e). Single-Wire Telephones for Mines Rescue. *CREGJ* 27, 7-11.
- (1998). Telephones for Mines Rescue. *Journal of the British Association of Public Safety Communications Officers (BAPCO)* 3(5), 6-8.
- Hill, D. and J. R. Wait (1973). Subsurface Electromagnetic Fields of a Grounded Cable of Finite Length. *Canadian Journal of Physics* 51, 1534-1540.
- Jennison, R. C. (1994a). G2AJV Toroidal Antenna: Part 1. *Radio Communication*, April 1994, 13-15.
- (1994b). G2AJV Toroidal Antenna: Part 2. *Radio Communication*, May 1994, 66-68.
- (1995). The G2AJV Antenna and Maxwell's Displacement Current. *Communications Quarterly*, Summer 1995, 23-34.
- Kraichman, M. (1962). Impedance of a Circular Loop in an Infinite Conducting Medium. *J. of Research of the Nat. Bureau of Standards - D. Radio Propagation* 66D(4), 499-503.
- Kraus, J. D. (1950). *Antennas*. New York: McGraw Hill.
- Lippold, R. (2000). Computer Simulation of a Grounded Wire Antenna. *CREGJ* 42, 5-11.
- Ramo, S., J. R. Whinnery and T. Van Duzer (1984). *Fields and Waves in Communication Electronics*. New York: John Wiley. (2nd Edition).
- Smith, F. L., Ed. (1940). *Radio Designer's Handbook*. London: Wireless World.
- Smith, G. S. (1972a). Proximity Effect in Systems of Parallel Conductors. *Journal of Applied Physics* 43(5), 2196-2203.
- (1972b). Radiation Efficiency of Electrically Small Multiturn Loop Antennas. *IEEE Transactions on Antennas and Propagation* 20(5), 656-657.
- Smith, G. S. and L. N. An (1983). Loop Antennas for Directive Transmission into a Material Half Space. *Radio Science* 18(5), 664-674.
- van Suchtelen, H. (1952). Ferroxcube Aerial Rods. *Electronic Application Bulletin* 13(6), 88-100.
- van Voorhies, K. L. and J. E. Smith (1992). Toroidal Antenna. *United States Patent* 5,442,369. USA, West Virginia University.
- Wait, J. R. (1995a). Excitation of a Conducting Half-Space by a Toroidal Coil. *IEEE Antennas and Propagation Magazine* 37(4), 72-74.
- (1995b). Excitation of a Conducting Half-Space by a Toroidal Coil: Correction. *IEEE Antennas and Propagation Magazine* 37(5), 43.
- Watt, A. (1967). *VLF Radio Engineering. [International Series in Electromagnetic Waves: 14]*. Oxford: Pergamon Press.
- Wolff, E. A. (1966). *Antenna Analysis*. New York: John Wiley & Sons.

5 Transmitter Design

Abstract: *The design of a transmitter is discussed with particular reference to the methods of driving the antenna from a power amplifier. It is shown that there are considerable differences between using an induction loop tuned to resonance and an untuned loop and that, for wideband systems with an untuned loop, there is a reduction in the magnetic moment we can achieve, although some benefits are also possible. It is shown that a wideband system should ideally utilise digital data or, at the least, a class-D 'digital' driver. Several areas for further study are highlighted.*

The previous chapter discussed some possible antenna systems, based on a classification by H , E , or J -field, and looked at aspects common to both transmitting and receiving antennas. Previously, we analysed the propagation by considering the fields from an induction loop H -field transmitter.

In this chapter we will concentrate on the design of induction loop transmitter antennas and the methods of driving them. We will use the concept of specific aperture, discussed earlier, to compare the performance of different systems. It will be seen that the ideal antenna depends on the intended use of the equipment and that a communications system may benefit from utilising different antennas depending on circumstances.

We will mention modulation methods and channel-coding only briefly in this thesis although they are clearly important in the design of a communications system.

5.1 Antenna Comparisons

5.1.1 Tuned v. Untuned Antennas

Induction loop antennas are, by their nature, highly inductive. The magnetic moment is proportional to the current in the antenna, but the inductance means that to maintain even a small current can require a high driving voltage. In addition, power losses in the driver can be severe when driving an inductive load.

For this reason it is usual to tune an induction loop to resonance with a series capacitor when, to a good approximation, it will present a resistive load to the driver. Unfortunately, the antenna may have a very high Q -factor when it is tuned in this way. Even if the bandwidth is high

enough for speech, it will almost certainly be too low for any wideband system we may devise. Additionally, the high Q -factor may place a large voltage stress on the tuning capacitor leading to unwanted power dissipation and parametric distortion.

5.1.2 Air-core v. Ferrite-core Antennas

In the previous chapter we gave a brief comparison between typical air-cored loops and ferrite-cored solenoids. We demonstrated that it was not possible to draw any definite conclusions about the relative merits without having a specific communications system in mind. For example, if the size of the antenna is not critical then a large air-core loop is preferred. We saw that a ferrite-cored transmitter antenna was likely to be small, but heavy, and that its performance was limited by the properties of the material.

5.1.3 Comparison by Power Dissipation

In chapter 3 we derived a figure of merit – the *specific aperture* – by which to compare antennas. The magnetic moment of the transmitter can be described in terms of the specific aperture and the power dissipation in the antenna winding. This is a more fundamental and practically useful parameter than, for example, the winding current or number of turns. In the derivations that follow, we relate the loop *power dissipation* to the characteristics of the power amplifier.

5.2 Driving a Tuned Antenna

For a narrow-band speech or beacon system we would usually tune the transmitter loop to resonance with a series capacitor, causing it to present a mainly resistive load to the power amplifier (PA). In this section we will derive an expression for the way this resistive load is matched to the output resistance of the PA and we will analyse the stress on the tuning capacitor. We will consider only the transmission of a carrier signal – a discussion of the efficiency of the power amplifier at frequencies other than the resonant frequency of the antenna will be deferred to section §5.3.1.1.

5.2.1 Matching the Antenna to the Amplifier

We will first consider the situation where we drive a tuned antenna from a conventional class B/C amplifier. *Figure 5-1* shows such an arrangement in an idealised form. Feedback reduces the high distortion that would otherwise occur, but distortion may not be a problem if we are driving a tuned antenna. With a high gain pre-amp and adequate feedback, the top transistor will conduct for half a cycle and the bottom transistor on for the other half cycle. We

will assume, to begin with, that the load can be considered to be purely resistive, which is a good approximation if we are driving a tuned antenna with a narrow-band signal.

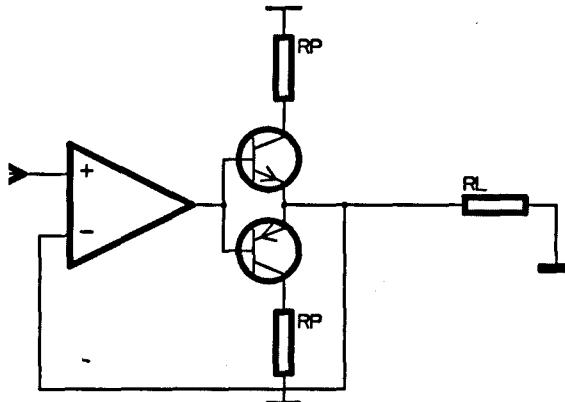


Figure 5-1 : Class-B/C push-pull power amplifier driving a resistive load

When the load is purely resistive, we do not need to know the details of the driver, and we can analyse the circuit as an idealised power source with source resistance R_p feeding the load resistance R_l . It is straightforward¹ to show that the efficiency of the power transfer from amplifier to load is

$$\frac{P_l}{P_l + P_p} = \frac{\pi}{4} g \frac{1}{(1+1/\eta^2)} \quad (5-1)$$

where P_l is the power dissipated in the load, P_p is the power dissipated in the amplifier, g is the gain of the PA expressed as a fraction of the maximum possible before clipping in the load, and η is the matching coefficient, which describes the relationship between the load resistance R_l and the amplifier output resistance R_p and which can be adjusted by inserting a matching transformer. We define the matching coefficient as

$$\eta = \sqrt{\frac{R_l}{R_p}} \quad (5-2)$$

In similar fashion, the absolute output power can be written as

$$P_l = \frac{1}{2} g^2 \frac{U_p^2}{R_p} \frac{1}{(\eta+1/\eta)^2} \quad (5-3)$$

where U_p is the PA rail voltage and R_p is the effective PA output resistance.

1 Instead of showing the derivation of (5-1) we will simply note that the configuration is a special case of the analysis for an inductive load, which will be given later in this chapter; and (5-1) can be obtained by setting $Q = 0$ in the equations set out in appendix A3.

It is important to realise that R_p is the source impedance of the power amplifier and not its output impedance. The output impedance of the idealised circuit of *Figure 5-1*, above, (i.e. the Thévenin equivalent circuit) is actually zero, due to the voltage feedback.

The η term in (5-3) indicates a classic ‘matching’ condition, and the expression reaches a maximum when $\eta = 1$. However, when this is the case, (5-1) indicates the well-known result that the efficiency of the power transfer is only 50% of the maximum possible. As η increases, the efficiency increases but the amount of power we can transfer decreases. Our system, being little more than an audio device (i.e. circuit and antenna dimensions much smaller than a wavelength) does not require us to control power reflection by means of impedance matching. Neither do we usually need to match for maximum power transfer because our power source is essentially a limitless source of energy². A possible compromise between efficiency and power capability would be to allow $\eta = 3$, which would allow us to transfer 36% of the maximum possible power at an efficiency of 90%. Alternatively, $\eta = 10$ would allow us to transfer 4% of maximum power at 99% efficiency. In most situations $\eta = 10$ would be a good design aim.

5.2.2 A Figure of Merit for Bandwidth Comparisons

It was shown in chapter 3 that the bandwidth of a tuned antenna does not depend on the number of turns but, more fundamentally, on the mass and the radius (or the length, for the case of a ferrite rod antenna). To obtain a low enough *Q*-factor for speech, an antenna with a large specific aperture must also be physically large, and of small mass. In contrast, radiolocation (beacon) transmitters and low-rate data transmitters can be small and heavy. Some comparisons between antennas were given in the previous chapter.

Since the magnetic moment for an air-cored loop antenna is $m_d = \Phi\sqrt{P}$ and the *Q*-factor is proportional to Φ^2/r^3 we can see that

$$\frac{m_d^2 B}{r^3} \propto P_l \quad (5-4)$$

and, consequently, for antennas with the same power dissipation, the quantity on the left is a constant, and is independent of specific aperture. This provides a figure of merit for comparing antennas but, as it stands, it is not particularly useful unless we also have a means of altering the bandwidth of an antenna independently of its other parameters.

2. For similar reasons, the 8Ω resistance of domestic hi-fi loudspeakers is not matched to the sub-ohm source impedance of the audio PA.

5.2.2.1 Equalisation

One method of adjusting the transmitter bandwidth is to apply equalisation to the signal; that is, to filter the signal so that the frequencies either side of the centre frequency are boosted by an appropriate amount. We could envisage operating a high- Q , low bandwidth antenna with a reduced output ($g < 1$) at resonance, and contriving to increase the PA output (g) for off-centre frequencies, such that g reached unity at the very edges of the spectrum (i.e. as $f \rightarrow 0$ and $f \rightarrow \infty$). This can be done in such a way that the response of the antenna exactly matches that of a lower- Q , wider bandwidth antenna operating at a lower power. We have shown (Gibson, 1995a) that, in this situation, (5-4) still applies, with B the equalised bandwidth and m_d the magnetic moment at resonance. The arguments are complex, and we have not performed a formal analysis. Additionally, in practice, we would probably wish to apply a simpler ‘flat-top’ equalisation, using current feedback from the antenna. Nevertheless, (5-4) can be considered a suitable figure of merit relating bandwidth to magnetic moment. This area of work is clearly a suitable subject for further analysis.

5.2.2.2 Tuning Capacitor

The tuning capacitor is a crucial item, since it will be handling a high current at a high frequency, and will have a large reactive voltage across it. For this reason, conventional induction radios with tuned and physically small antennas do not always meet the intrinsic safety standards required for mine rescue systems. The tuning capacitor may turn out to be the single most ‘lossy’ part of the antenna, so it is an important consideration.

If we can represent the complex impedance of a reactive component as $R + jX$ then the phase of the voltage relative to the current can be represented as δ , with

$$\tan \delta = R/X \quad (5-5)$$

The quantity $\tan \delta$ is known as the *dissipation factor* and is the reciprocal of Q -factor. The angle δ is known as the *loss angle* and we can also define $\sin \delta$ as the *power factor*. It is easy to show that in a tuned circuit consisting of an inductor and capacitor, both with loss resistances, that the overall dissipation factor is simply the sum of the two individual dissipation factors or, in terms of Q ,

$$\frac{1}{Q_{total}} = \frac{1}{Q_C} + \frac{1}{Q_L} \quad (5-6)$$

and the power dissipated in the inductor will be

$$P_i = \frac{Q_{total}}{Q_L} = \frac{Q_C}{Q_{total}} \quad (5-7)$$

This is significant because it shows that a poor quality capacitor, with a $\tan \delta$ of, say 0.1 at the operating frequency, will easily dominate the overall Q -factor and lead to a low antenna efficiency.

A review of suitable low-loss capacitors was given in (Gibson, 1995b) and concluded that metallised polypropylene capacitors were the most suitable for our purposes, with voltage ratings up to 2000V and $\tan \delta$ as low as 0.001 at 100kHz. Another possibility would be high-performance ceramic capacitors with dielectrics such as 'Z5U', which are used in switched-mode power supplies. However, although these have a good power rating, their dissipation factor is higher than polypropylene. Capacitor manufacturers can easily tailor devices to customer specifications and this may be worth considering for commercial equipment.

We must also consider the voltage and current stresses on the capacitor. We will need to limit the reactive voltage to a safe value, but even at moderately high values we may see effects due to parametric distortion. For many dielectrics, capacitance is not a completely linear function of voltage and this has given rise to some observable effects in cave radio equipment.

At resonance, the voltage across either reactive element is $I X$ which we can write in terms of the inductor parameters, as $Q_i I R_i$. Using (4-18) we can write the reactive voltage as

$$U_X = \omega \pi K_Q \frac{N}{r} \Phi \sqrt{P_i} \quad (5-8)$$

which demonstrates that, for a given magnetic moment, the reactive voltage is minimised by choosing a smaller number of turns or a larger radius. We already know, from the discussion in §3.2.6 that the bandwidth is unaffected by the number of turns. Writing this expression in terms of $m_d = \Phi \sqrt{P}$ and multiplying both sides by $m_d = NIA$ results in

$$m_d = \sqrt{\frac{r^3 U_X I}{\omega K_Q}} \quad (5-9)$$

which is important because it indicates that if the reactive voltage and antenna current are constrained by practical component specifications then the magnetic moment depends on these specifications and not on specific aperture. In other words, if we construct an antenna with a large specific aperture we can pass only a small amount of power into it before we reach the design limit of voltage or current. It is interesting to note the similarity between (5-9) and the equation for the magnetic moment of an untuned antenna given later in (5-16).

Given this equation as a design rule, we are free to choose Φ and P as we like, although, having done this, the parameters N and R will be fixed because we have also specified the current I in the above equation.

5.2.2.3 Frequency Drift

In addition to leading to transmitter inefficiency, the result of a poor dissipation factor is that the capacitor will rise in temperature. If it has a poor temperature coefficient then the antenna is likely to drift off-tune. If the capacitor heats from 20°C to 70°C then a temperature coefficient as low as 200ppm will result in a drift of 1%, corresponding to 1kHz at 100kHz. We suspect that the tuning drift, in addition to dissipation losses, could account for the sometimes-poor performance of amateur-built induction radio equipment.

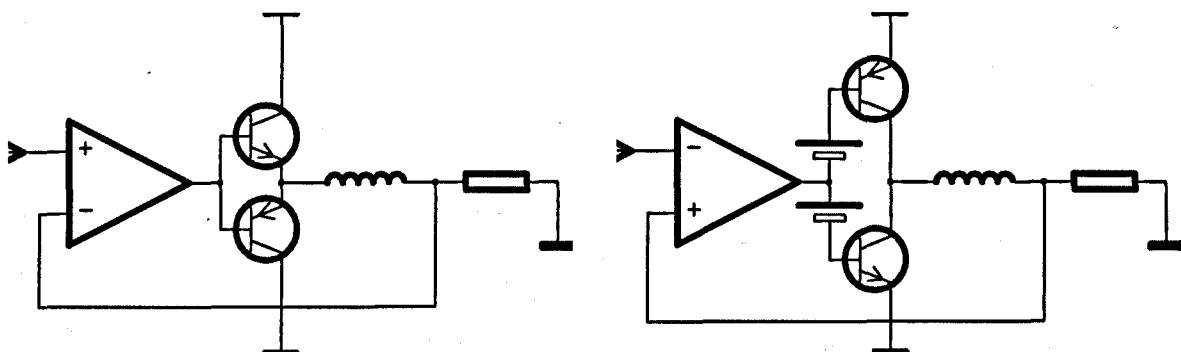
One method of alleviating this problem would be to allow the operating frequency to be automatically adjustable as the resonant frequency of the antenna drifted. This would require a compatible receiver design using, for example, an untuned antenna, with the selectivity provided by a phase-locked loop or digital signal processing.

5.3 Driving an Untuned Antenna

Despite the inefficiency, there are times when we might be required to use an untuned antenna. An obvious case is when we wish to transmit a wideband signal, but we may also decide to use an untuned antenna for a spot-frequency beacon because it would avoid problems due to parametric drift and voltage stress in the tuning capacitor. In this section we will derive expressions for the efficiency of an untuned antenna and the power losses in the driver.

5.3.1 Analogue Drivers

We will first consider the problem of driving an untuned antenna from a conventional push-pull or complementary amplifier as shown, in outline form, in *Figure 5-2* which is powered from positive and negative supplies, with the load grounded at roughly the mid-rail point. With adequate feedback these configurations will perform in a similar fashion to a class-B amplifier – that is, the ‘top’ transistor will conduct for half a cycle, controlling the supply of positive-going current to the load, and the ‘bottom’ transistor will conduct for the other half cycle sinking the negative-going current.



*Figure 5-2 : Class-B/C amplifier driving an inductive load with current feedback.
Left: push-pull configuration. Right: Complementary configuration.*

The circuit is shown with current feedback from a sense resistor, but it could operate equally well using voltage feedback from the common-emitter output node. In practice, a current-feedback configuration may require additional compensation in the feedback path, but that need not concern us here³. We could also use this configuration as a digital driver without feedback; however, a computer simulation is easier to interpret if there is a feedback path present.

A detailed discussion and analysis is given in appendix A3. We will summarise the results by noting that the matching coefficient is not as crucial as it was for the case of a tuned antenna, and that for $\eta \geq 1$ (sic) and $Q \gg 1$ the equations reduce to

$$\frac{P_l}{P_l + P_p} \approx \frac{\pi}{4} g \frac{1}{Q} \quad (5-10)$$

and

$$P_l \approx \frac{1}{2} g^2 \frac{U_p^2}{R_l} \frac{1}{Q^2} \quad (5-11)$$

so, for a high- Q antenna, we can achieve only a low power level due to the presence of the $1/Q^2$ factor in (5-11). This *in itself* may not be a problem, but the *efficiency* is modified by the $1/Q$ term in (5-10) and this is a cause for concern.

5.3.1.1 Driving a Partially-tuned Antenna

- The result, that the efficiency of an PA is reduced when driving an inductive load, can be extended to the situation where the load is a resonant circuit, but is operated off-tune, so that it presents a partially reactive load. The equations are presented in appendix A3.3 and are necessary for the analysis of a tuned antenna that is transmitting a signal in a finite bandwidth. They are also useful for investigations of equalisation, and of drift in a tuned antenna, but we shall not pursue this topic further at the present time.

5.3.2 Digital Drivers

We could expect a digital output stage to be more efficient because it is either fully on or fully off. This is the case for a complementary driver, but it is not so for the push-pull configuration for reasons explained in appendix A3.2.2. In summary, the load power, assuming $Q \gg 1$, is

$$P_l = \frac{U_p^2}{R_l} \frac{\pi^2}{12} \frac{1}{Q^2} \quad (5-12)$$

³ A circuit analysis shows that current feedback from a floating inductive load can result in an instability that requires an RC compensation network. Alternatively, if the sense resistor is inductive this may aid stability. If the sense resistor is floating, with differential feedback, and the inductive load is grounded, then the network is often stable without additional compensation.

For the complementary output stage, with a 1:1 duty cycle, we can only operate it at full power, so the g term does not appear in the expressions. The amplifier power dissipation is R_p / R_l times the above expression, and so the ratio of the two, from which we can obtain the efficiency, is

$$\frac{P_l}{P_p} = \frac{R_l}{R_p} = \eta^2 \Rightarrow \frac{P_l}{P_l + P_p} = \frac{1}{1 + 1/\eta^2} \quad (5-13)$$

For the push-pull configuration, assuming $\eta \geq 1$ and $Q \gg 1$, the efficiency can be written as

$$\frac{P_l}{P_p} = g \frac{\pi}{6} \frac{1}{Q} \quad (5-14)$$

A switched-mode driver using a complementary output stage is therefore clearly the best option for transmitting a wideband digital signal. In a digital system, we may require to transmit a PWM signal and so the form of the complementary driver would be that of a re-circulating H bridge, as used in a standard ‘chopping’ motor-drive circuit, for example. *Figure 5-3*, below, outlines a half H-bridge driver configuration that we could also use.

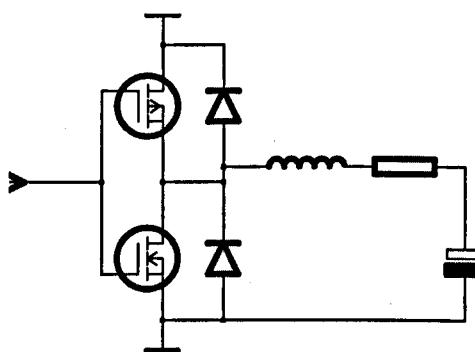


Figure 5-3 : Half H-bridge complementary output for digital amplifier.
The diodes provide a low-loss reverse current path and, in this example,
the load floats mid-rail due to the capacitor.

5.3.3 A Figure of Merit for Untuned Antennas

The salient points from the previous two sections are that the available power in an untuned antenna is proportional to $1/Q^2$ and that the efficiency is proportional to $1/Q$ for analogue signals but, for digital signals, the efficiency is related only to the matching coefficient. To be efficient, a wideband system must therefore use digital signals. If we were to consider such a system in detail then it would be appropriate to look at the efficiency of a PWM driver, rather than the simple 1:1 duty-cycle square wave that we considered above.

Merely knowing that the efficiency is related to Q is not sufficient to enable us to design an antenna, for which purpose a ‘figure of merit’ similar to the specific aperture Φ that was derived in chapter 3, is also needed. We will now derive some suitable figures of merit.

5.3.3.1 Analogue Drivers

We saw, in chapter 3, that the Q -factor of a loop was proportional to Φ^2/r^3 . To decrease Q we have therefore to increase the radius of the loop, r , or to decrease the specific aperture. Doing the latter will result in a lower magnetic moment but an increased efficiency in the driver. These two effects cancel out and the field strength becomes dependent only on the radius of the antenna (i.e. not on its specific aperture). The total power drawn from the power source is $P_e = P_l + P_p$ so we can write (5-10) as

$$P_l = \frac{\pi}{4} g \frac{1}{Q} P_e \quad (5-15)$$

We can substitute this in the expression for specific aperture (3-5) and use the expression for Q in (3-18) to obtain

$$m_d = \sqrt{\frac{1}{4} g \pi} \cdot \sqrt{\frac{r^3 P_e}{\omega K_Q}} \quad (5-16)$$

where K_Q is as defined in §3.2.6. This expression is valid for $\eta \geq 1$ and $Q \gg 1$. We can note the corresponding expression for a tuned antenna, which is

$$m_d = \Phi \sqrt{\frac{1}{4} g \pi} \cdot \sqrt{P_e} \quad (5-17)$$

and the expression for the magnetic moment as a function of the reactive power in the load, given in (5-9).

5.3.3.2 Digital Drivers

The above expression is only valid for analogue power amplifiers, because the efficiency of a digital output stage does not contain a Q term. However, knowing that the power *delivered* to the loop antenna from a digital driver (5-12) *does* contain a Q term, we can combine equations to give

$$\left. \begin{aligned} P_l &= \frac{U_p^2}{R_l} \frac{\pi^2}{12} \frac{1}{Q^2} \\ Q &= \frac{K_Q}{r^3} \Phi^2 \\ m_d &= \Phi \sqrt{P_l} \end{aligned} \right\} \Rightarrow m_d = \frac{r^3}{\omega K_Q \Phi} \sqrt{\frac{U_p^2}{R_l} \frac{\pi^2}{12}}, \quad Q \gg 1 \quad (5-18)$$

which suggests that a lower specific aperture gives rise to a larger moment. This is only true, of course, provided that Q remains $\gg 1$, because $\Phi \rightarrow 0$ does not, of course, imply $m_d \rightarrow \infty$. This result is difficult to interpret – the limitation on Q arises because, with a digital driver, we are assuming that we want the current to ramp linearly with time, but this need not necessarily be

the case. This expression is probably best investigated by using a spreadsheet to examine the effect of varying the parameters.

5.3.4 Comparisons & Remarks

5.3.4.1 Analogue Signals

For a tuned antenna we have to consider the matching and the specific aperture but, for an untuned antenna with an analogue PA, we need consider only the radius of the antenna, as expressed in the figure of merit in (5-16). The magnetic moment is independent of the mass of the antenna and the number of turns, except that if the mass is very low then the resistance will be high, and there will come a point when the matching requirement $\eta \geq 1$ no longer holds. This, however, simply means that we need to use the full expression for the efficiency, and not the approximation of (5-10).

Equation (5-10) and its full expression (A3-8) in appendix A3.1.1 show that a tuned antenna will always be more efficient unless it is badly matched (the matching being more significant for a tuned antenna).

However, there are some attractions to using an untuned antenna. For example, if we make it of a single turn of wire, then losses due to self-capacitance and the proximity effect will be non-existent. In some cases, therefore, although an untuned antenna is *theoretically* less efficient than a tuned antenna, it may prove as good, or better in practice. This is an important topic for future work.

5.3.4.2 Digital Signals

We saw that for digital signals a complementary driver could approach a high efficiency, unlike a push-pull driver for which the efficiency included a $1/Q$ term. A figure of merit (5-12) for a digital driver suggests that we can achieve a higher moment by using an antenna with a *lower* specific aperture. This is only true, of course, provided that the *Q*-factor of the antenna remains high.

5.4 Damped Antennas

We have seen that using an untuned antenna results in a lower power transfer efficiency, although it can also result in a wider bandwidth. If we wished deliberately to increase the bandwidth we might decide to construct a smaller antenna, but this could lead to problems with antenna matching. We could also damp an antenna by deliberately adding a series resistance. In addition, of course, this may aid the matching of the antenna to the PA.

5.4.1 Damping a Tuned Antenna

A tuned antenna may have too low a bandwidth for speech. Suppose we increase the resistance of the tuned circuit by ζ times by adding a damping resistor. If $Q \gg 1$, and remains so, then the Q -factor will go down by ζ times and (5-3) indicates that the power we can deliver will fall by ζ times. In addition, the load power is being dissipated in the additional resistance, as well as in the antenna, and the proportion in the antenna is $1/\zeta$ of the total. The result is that the load power reduces by ζ^2 times and the magnetic moment by ζ times. We can show that the overall effect is the same as if the additional resistance were obtained by winding the antenna with thinner wire.

Conclusion: damping by adding a series resistance increases the bandwidth, and might aid the matching, but the effect is the same as reducing the specific aperture and is therefore not desirable.

5.4.2 Damping an Untuned Antenna

5.4.2.1 Analogue Driver

Suppose we increase the load resistance by ζ times by adding a resistor in series with the transmitter loop. If $Q \gg 1$, and remains so, then the Q -factor will reduce by ζ times and, (5-11) indicates that the power we can deliver will rise by ζ times (because we have also increased the R_L term by ζ times). However, the load power is being dissipated in the additional resistance, as well as in the antenna, and the proportion in the antenna is $1/\zeta$ of the total. The result is that the overall magnetic moment of the antenna is unchanged. Additionally, by decreasing Q we have increased the efficiency of the power amplifier by ζ times, but we require to draw ζ times more power.

Conclusion: damping by adding a series resistance increases the bandwidth, and might aid the matching. There is no change to the power drawn from the energy source or the magnetic moment of the transmitter.

Another method of damping the antenna would be to use thinner wire. Suppose we decrease the wire c.s.a. by ζ times whilst keeping the other parameters unaltered. The resistance will increase by ζ times, as will the bandwidth; and the specific aperture will reduce by $\sqrt{\zeta}$. As before, we can deliver ζ times more power but, this time, the power is all dissipated in the loop. However, because the specific aperture has reduced, the magnetic moment remains unaltered. As before, we have increased the efficiency of the power transfer, but we will draw more power. This result therefore confirms that the moment is independent of specific aperture.

This observation is notable because it can be used to explain the effect of skin resistance. Skin and proximity effects change the a.c. resistance of the wire, thus altering either its apparent conductivity or its apparent mass – either way, they alter the specific aperture. But the magnetic moment does not depend on specific aperture for an untuned antenna, so we can deduce that the skin effect is not detrimental.

5.4.2.2 *Digital Driver*

The above arguments apply to a digital driver, except that the PA efficiency is governed by the matching coefficient and does not benefit from the decrease in *Q*-factor. This means that adding a damping resistor simply causes us to waste power.

5.5 Modulation Methods

In this thesis we will not have much to say about modulation methods, although clearly modulation and channel coding decision are an important part of a system design.

5.5.1 Analogue Methods

5.5.1.1 *Single Sideband*

For speech systems, many induction systems (both for cave and mine operation) utilise *single-sideband* (SSB) operation because this is an efficient use of the transmitter power, and because adjacent channel interference often dictates the use of a narrow bandwidth. There are several methods of generating and demodulating SSB signals. The method of choice for demodulation depends on whether the receiver is a conventional analogue device or DSP-based. For analogue receivers the so-called phasing method has proved to be popular. This relies on certain characteristics of all-pass filter networks, for which original design information is not readily available. The definitive work is probably that written by the author, (Gibson, 1996a; b). For digital receivers a method due to (Turner, 1973; Weaver, 1956) is preferred, as outlined in (Gibson, 2002). Other methods of SSB generation and detection, such as the ‘polyphase’ filter are outdated and to be deprecated.

5.5.1.2 *Carrier Methods (AM, FM)*

In general, carrier-borne systems such as *amplitude modulation* (AM) and *narrow-band frequency modulation* (NBFM) are to be avoided because of the large amount of power needed to generate the carrier. However, when transmitting an SSB signal there is no operational reason why a vestigial carrier, and perhaps a vestigial sideband, should not be transmitted if it would be costly to suppress them.

5.5.1.3 Double Sideband, Suppressed Carrier

The other possible analogue method to consider is that of *double-sideband suppressed carrier* (DSBSC). For a given transmitted power, a DSBSC system confers the same SNR as an SSB system, provided that it is demodulated correctly using a reconstructed, synchronised carrier. If, however, the transmitter is peak-power limited then we can transmit only half the total power that is possible for an SSB system, so the SNR is down by 3dB.

DSBSC systems are usually not considered for use with portable communications because, although the transmitter is much simpler than an SSB transmitter, the receiver requires good phase synchronism, needing additional circuitry. This is, however, straightforward to achieve using one of several well-known methods. For data transmission, it may be advantageous to consider a DSBSC system, because one of the common methods of demodulating SSB introduces a large frequency-dependent phase shift. This is unimportant for speech, but could make it impossible to operate a data modem.

5.5.2 Digital Methods

For its simplicity, narrowband digital data transmission can be achieved with a simple phase-shift keying (PSK) method. Differential PSK may be preferred; and this could be binary, or quadrature. Apart from noting that the appropriate measure of the effect of noise is the bit error rate (BER) that is dependent on the ratio of signal energy to noise power density (E / N_0) we do not intend to discuss digital methods at this time.

5.6 Concluding Remarks

For a transmitter antenna the main design choices are that of a tuned v. untuned antenna and a ferrite v. air-cored antenna. We saw, in the previous chapter, that ferrite cored antennas had limited use as transmitter antennas and, in this chapter, we concentrated on the comparison between tuned and untuned antennas. The results are summarised in *Table 5-1* below.

We discussed modulation methods briefly, noting that single-sideband (SSB) was a good choice for a conventional speech system, but that data transmission over an SSB speech link might be subject to errors because of the unfavourable phase shift.

For any communications system, its performance depends ultimately on the characteristics of the transmitter system – including the data and channel coding and the available power. Often, the system will benefit from the transmitter antenna being as large as possible. Through-rock communications systems that require only a ‘down-link’ transmission can often utilise a physically large surface transmitter. However, for an up-link transmission, the antenna might be restricted in mass or size. Not only does this mean that the transmitter and receiver antenna may

have different designs, but the up-link and down-link may also require different designs. The information presented in this chapter can be used in the design of a transmitter antenna that is tailored to a particular communications system. In particular we can make use of the analysis of an untuned antenna in the design of a wideband antenna suitable for channel sounding. We will discuss channel sounding in chapter 7, but prior to that, in the next chapter, we will undertake a detailed analysis of the receiver antenna.

Antenna	Characteristics	Equation
Tuned antenna, not equalised	<ul style="list-style-type: none"> ▪ High m_d for given power consumption; (5-17) ▪ Narrow-band ▪ Mass of antenna (Φ) affects performance ▪ Antenna can be impedance-matched to PA ▪ Performance affected by skin and proximity effects ▪ If tuning capacitor is a design constraint then a performance constraint is (5-9) 	$m_d = \Phi \sqrt{\frac{1}{4} g\pi} \cdot \sqrt{P_e}$ (5-17)
§5.2.1		$m_d = \sqrt{\frac{r^3 U_x I}{\omega K_Q}}$ (5-9)
Tuned antenna, equalised	<ul style="list-style-type: none"> ▪ m_d / \sqrt{P} ratio depends on chosen bandwidth; (5-4) ▪ Bandwidth is adjustable ▪ Mass of antenna (Φ) does not affect performance ▪ Matching: <i>as for non-equalised antenna</i> ▪ Performance is not affected by skin or proximity effects ▪ Tuning capacitor: <i>as for non-equalised antenna</i> 	$\frac{m_d^2 B}{r^3} \propto P_l$ (5-4)
§5.2.2		
Untuned antenna, class B/C driver	<ul style="list-style-type: none"> ▪ Low m_d for given power consumption; (5-16) ▪ Power efficiency is proportional to $1/Q$ and absolute power level in antenna proportional to $1/Q^2$ ▪ Wide-band, can be adjusted by damping without affecting performance ▪ Mass of antenna (Φ) does not affect performance ▪ Matching is not an issue ▪ Performance is not affected by skin or proximity effects 	$m_d = \sqrt{\frac{1}{4} g\pi} \cdot \sqrt{\frac{r^3 P_e}{\omega K_Q}}$ (5-16)
§5.3.1, §5.3.3.1		
Untuned antenna, class D driver	<ul style="list-style-type: none"> ▪ Medium m_d for given power consumption ▪ Power efficiency is not restricted by Q-factor, but absolute power in antenna proportional to $1/Q^2$ ▪ Wide-band ▪ Mass of antenna (Φ) affects performance ▪ Matching may be an issue ▪ Performance affected by skin and proximity effects 	$m_d = \frac{r^3}{\omega K_Q \Phi} \sqrt{\frac{U_p^2 \pi^2}{R_i}} \sqrt{\frac{12}{1}}$ (5-18)
§5.3.2, §5.3.3.2		

Table 5-1 : Comparison of tuned v. untuned antennas.

5.7 References⁴

- Gibson, D. (1995a). Bandwidth of Tuned Antenna does not Affect Performance. *CREGJ* 21, 23-26.
- (1995b). Losses in Tuning Capacitors. *CREGJ* 19(12-14).
- (1996a). Designing an SSB Out-Phaser (Part 1). *Electronics World* 102(1721), 306-310.
- (1996b). Designing an SSB Out-Phaser (Part 2). *Electronics World* 102(1722), 392-394.
- (2002). Cave Radio Notebook 48: A DSP Implementation of the Turner / Weaver SSB Modulator. *CREGJ* 47, 18,24.
- Turner, A. J. (1973). Single-Sideband Suppressed Carrier Generation - Modification of the 'Third Method' Made Possible by the Use of Integrated Circuits. *Wireless World* 79(1455), 453-455.
- Weaver, D. K. (1956). A Third Method for the Generation and Detection of Single-Sideband Signals. *Proc. IRE* 44(12), 1703-1705.

⁴ References to *CREGJ* are to the *Cave Radio & Electronics Group Journal*; a non-peer-reviewed low-circulation magazine published by a special interest group of the *British Cave Research Association*, bcra.org.uk. The CREG journal is filed at the British Library as ISSN 1361-4800.

6 Noise and Receiver Design

Abstract: The design of a receiver antenna is discussed with particular reference to its noise factor. Two contributions to the noise factor are the atmospheric (external) noise, and the amplifier circuit (internal) noise. The likely level of atmospheric noise is discussed, together with the concept of a 'maximum useful size' of the receiver antenna. For the amplifier noise, it is shown how this is related to the impedance of the antenna; and that this can be optimised using the concept of noise-matching. It is shown that, for wideband systems with an untuned loop, there is a considerable reduction in the noise factor caused by the high reactance of the antenna. It is further shown that this can be countered by series-tuning the antenna and damping it to widen the bandwidth. Alternatively, a physically larger antenna will have a wider bandwidth. A brief discussion of suitable operational amplifiers is given, and some typical low-noise devices are highlighted. Finally, methods of mitigating atmospheric noise and interference signals are discussed, including shielding, nulling and cancellation strategies.

Through-rock communications systems may have to work with a very low signal level combined with a high level of background noise, co-channel and adjacent channel interference, so the design of the receiver antenna – and its noise performance in particular – is important. We have seen how the concept of specific aperture (Φ), developed in previous chapters, allowed us to relate the magnetic moment of the transmitter to the power dissipated in its winding. We saw that the same figure of merit could be applied to a receiving antenna, for which it could represent the *sensitivity* of the antenna thus demonstrating that the receiver SNR does not depend on the number of turns of wire, nor on its diameter, but only on the more fundamental properties such as the mass of the wire, its conductivity and density. As with the transmitter design, this figure of merit shows that considerations such as the number of turns and the wire diameter are not of primary importance. For a ferrite-cored solenoid, we replace Φ by $\mu_r \Phi$.

We also saw, in chapter 3, that we could incorporate the atmospheric noise in a 'figure of merit' equation, and could thereby conclude that there was a maximum useful size for the receiver antenna. In this chapter we will further develop this concept, by including the receiver pre-amp noise in our calculations. Fundamentally, it can be appreciated that the antenna should be large enough for the received signal to swamp both the circuit noise and the thermal noise of the antenna but that, once this size has been exceeded, there is no purpose to making this

antenna any larger. Clearly, for the design of a communications system, it is helpful to know this maximum useful size.

6.1 Fundamentals of Antenna Design

6.1.1 Antenna Comparisons

6.1.1.1 Tuned v. Untuned Antennas

For a transmitting antenna we saw, in the previous chapter, that if we wished to transmit a wideband signal we would have to use an untuned antenna, which was far less efficient than a tuned antenna. For a receiving antenna we can make the same choice between using it tuned or untuned; the main difficulty being that an untuned antenna presents a high-impedance inductive source to the receiver pre-amp, which can lead to a high noise voltage if the receiver has a significant input bias current.

An untuned induction loop antenna acts like a low-pass filter but we may wish to operate above this cut-off frequency, so that the inductor acts as an integrator. Since the induced voltage is proportional to dB/dt , this would result in an output signal that was independent of frequency. For channel sounding, however, the transmitter antenna acts to integrate the wideband binary sequence we wish to use (as discussed in the following chapters) and so we may require the receiver simply to differentiate the signal – in which case we would wish to be operating in the ‘low pass’ region of the response. Clearly the design of the antenna depends very much on the purposes for which it is intended to be used.

6.1.1.2 Air-core v. Ferrite-core Antennas

In chapter 4 we discussed the relative merits of air-cored loops and ferrite-cored solenoids. Applying these results to the design of a receiver antenna is straightforward, with a long thin ferrite rod demonstrating the most efficient use of the high permeability of the material. Unfortunately, using a ferrite rod can result in a very high Q -factor and so the antenna may have to be used untuned, when its high inductance can lead to a high noise, as noted above.

6.1.2 Signal / Noise Ratio

6.1.2.1 A General Expression

During the discussion on specific aperture, in §3.3–3.4, we derived an expression for the signal /noise ratio of a receiver antenna, incorporating the thermal noise of the antenna and the atmospheric noise. Adding a matching transformer between the antenna and amplifier, with a step-up turns ratio of η , the signal elements become

a) Signal voltage $\omega\mu H \cdot \eta \cdot \Phi \sqrt{R}$ (6-1a)

b) Thermal noise $\sqrt{4kT_0 BR} \cdot \eta$ (b)

c) Atmospheric noise $\sqrt{4kT_0 BF_a R_r} \cdot \eta$ (c)

To these expressions¹ we will now add the noise of the receiver pre-amplifier. This can be expressed as an *equivalent input noise voltage* with a spectral density² of \hat{v}_n [V/ $\sqrt{\text{Hz}}$] and an *equivalent input noise current* with a spectral density of \hat{i}_n [A/ $\sqrt{\text{Hz}}$], which flows in the circuit resistance to produce a noise voltage. The additional noise contributions are...

d) Amplifier noise voltage $\hat{v}_n \sqrt{B}$ (d)

e) Amplifier noise current in antenna, R $\hat{i}_n R \cdot \eta^2 \cdot \sqrt{B}$ (e)

(This assumes the condition $R_r \ll R$).

f) Terms due to bias resistor

A ‘bias’ resistor R_b , connected to the amplifier input, will contribute a thermal noise not previously accounted for as well as a noise voltage due to the amplifier’s noise current.

In many circumstances, R_b can be discounted – we will do so here, in order to simplify the analysis, but there is a further discussion in appendix A4.1.

We stated that B was the noise-equivalent bandwidth of the antenna. However, by including it here, we have made the assumption that B is also the noise-equivalent bandwidth (NEB) of the amplifier. Although this may not be the case in practice, it is a convenient assumption on which to base the analysis. It is assumed that \hat{v}_n and \hat{i}_n are constant with frequency. We are also assuming, of course, that the atmospheric noise is Gaussian in nature. It is known that, below 30kHz, the noise becomes increasingly impulsive due to its origin in lightning strokes. The effect of the noise on communications can be considerably less than would be indicated by considering its r.m.s. level, provided that there is adequate signal processing at the receiver – e.g. see (Burley, Darnell, Prowse and Chandler, 1997). Consequently, the receiver antenna can be constructed to have a higher sensitivity than would be expected from a study of the r.m.s. noise level.

- 1 The symbols are as defined in §3.3 – that is, ω is frequency, μ is the permeability of the receiver antenna, Φ is the specific aperture, R is the resistance of the receiver loop, k is Boltzmann’s constant, T_0 is the ambient temperature – usually taken to be 290K, B is the noise-equivalent bandwidth of the system, F_a is the atmospheric noise temperature ratio (as derived from a magnetic loop antenna), and R_r is the antenna’s radiation resistance.
- 2 These spectral densities are *defined* as the input-referred r.m.s. noise voltage divided by \sqrt{B} . Note that spectral density – especially when derived using Fourier transforms is sometimes defined as a ‘two-sided’ value, i.e. using $\sqrt{2B}$.

Additionally, the above expressions assume that the antenna presents a purely resistive source. In practice the antenna will be reactive, or resonant, requiring a more detailed analysis, which we shall give in a later section.

6.1.2.2 Noise Factor

We define the *noise factor* as the ratio of the amplifier's input to output SNR. Each of the terms in (6-1) can be minimised by suitable design or operating conditions, except the thermal noise of the antenna, which we cannot avoid. Hence, it is appropriate to define the input SNR as the ratio of the signal to the antenna thermal noise, and to have the output SNR include the degradation of the amplifier *and* the contribution from the atmosphere. Therefore, writing the noise factor F_n as a power ratio, we obtain the output SNR as

$$\frac{S}{N} = \frac{\omega\mu H}{\sqrt{4kTB}} \Phi / \sqrt{F_n} \quad (6-2)$$

with

$$F_n = 1 + F_a \frac{R_r}{R} + \frac{1}{4kT} \left(\frac{v_n^2}{\eta^2 R} + i_n^2 \eta^2 R \right) \quad (6-3)$$

and so, making the substitutions of

$$\hat{P}_n = \hat{v}_n \cdot \hat{i}_n, \quad \text{and} \quad R_n = \hat{v}_n / \hat{i}_n. \quad (6-4)$$

and using the shorthand notation of $\|x\| \equiv x + 1/x$, we can write the noise factor as

$$F_n = 1 + (F_k \Phi)^2 + \frac{\hat{P}_n}{4kT} \left\| \frac{R_n}{\eta^2 R} \right\| \quad (6-5)$$

where \hat{P}_n is the *amplifier noise power spectral density*, R_n is the *noise resistance*³, and F_k is an 'atmospheric noise term' derived from F_a . The derivation was given in §3.4.1.1 and we will discuss it again in §6.2.5. The term containing R_n describes a classic 'matching' situation, and we can easily show that the term has a minimum value of 2 when

$$R_n = \eta^2 R \quad (6-6)$$

that is, when the amplifier's noise resistance is matched to the *apparent* source resistance of the antenna, allowing for the matching transformer. However, we must note well that this particular matching condition assumes that the antenna is purely a resistive source.

The above analysis assumes that the matching transformer is 'ideal' whereas, if used, this could have a significant effect on the frequency response, as described in (Paschal, 1988).

3 It is important to note that noise resistance, as defined here, is the ratio of noise voltage to noise current, and *not* – as it is in some texts – the value of an input resistor that would deliver the same thermal noise that the amplifier generates.

6.1.2.3 Designing using the Noise Factor

The expression for the noise factor, in (6-5) is significant because it demonstrates the conditions that may lead to one noise source being predominant. Depending on the antenna design, we may be able to minimise the amplifier noise, by noise-matching, to a value of $\hat{P}_n / 2kT$. Some modern op-amps have a sufficiently low noise power ($< 4\text{zW/Hz}$)⁴ that this term could be below 0.5, which makes it insignificant in comparison to the thermal noise, represented by the first term in (6-5). Similarly, as discussed in §3.4.1.1, a sufficiently large specific aperture will mean that the atmospheric noise will predominate. We will now consider the noise terms individually, in detail.

6.2 Atmospheric Noise

One of the purposes of our channel sounding experiments will be to measure atmospheric noise and interference for different antenna designs and communications strategies. Prior to this, it will be useful to have a rough indication of the likely amount of noise. We can classify the external noise sources as a) atmospheric, largely due to world-wide electrical storms; b) galactic, which is not significant at the frequencies we shall be considering; and c) man-made interference, which can exceed atmospheric noise above about 100kHz except at very quiet rural locations. The characteristics of these noise sources have been documented by the CCIR⁵, which is now part of the ITU⁶. The often-quoted report 322 (CCIR, 1968) has been superseded by ITU report 372 (ITU, 2001) which presents similar data, which we will use in the discussion that follows.

6.2.1 Atmospheric Noise Temperature Ratio

In §3.3.2 we noted that the atmospheric noise field could have a fictitious temperature T_a assigned to it, such that it delivered a noise power of kT_aB to the antenna's radiation resistance, where k is Boltzmann's constant and B is bandwidth⁷. Since many sources quote atmospheric noise as a temperature ratio (that is, as dB above kT_0) rather than as a noise field spectral density, it is appropriate to establish a relationship between the *atmospheric noise temperature ratio* F_a and the *magnetic equivalent noise field spectral density*⁸ \hat{H}_n . The atmospheric noise can be considered to be the sum of many individual field vectors that combine to give rise to a noise voltage in the receiver. We can therefore calculate the 'equivalent' electromagnetic field

4 The SI prefix *zepto*, symbol 'z', indicates a unit multiplier of 10^{-21} .

5 CCIR = Comité Consultatif International des Radiocommunications.

6 ITU = International Telecommunication Union.

7 Strictly speaking, the *noise-equivalent* bandwidth.

8 A circumflex over the symbol is used, in this thesis, to denote a spectral density.

that would, if it were incident on the antenna, give rise to the same noise voltage. However, as we will show in the discussion that follows, this ‘equivalent’ field is a fictitious construction that depends on the antenna design. Unfortunately, the noise temperature data quoted in the CCIR/ITU reports cannot reliably be converted to an equivalent magnetic noise field for two reasons.

- i) **Relationship between E and H fields.** The CCIR noise measurements were made using an electric antenna (specifically a short vertical monopole). Because the measured noise is the vector sum of many different noise sources it cannot be equated to a single electromagnetic wave, and so the E and H -fields are not necessarily related by the wave impedance Z_0 . Additionally, not all the noise detected by the electric antenna may be electromagnetic in origin, as there may be electrostatic contributions from man-made interference, especially at higher frequencies. Therefore, although we can derive a figure for the electric ‘equivalent noise field’ spectral density we cannot legitimately convert it to a magnetic ‘equivalent noise field’ by dividing by Z_0 .
- ii) **Antenna Orientation.** A horizontal loop antenna (vertical magnetic dipole, VMD), such as is common for sub-surface communications, will be most sensitive to a vertical magnetic field, but the corresponding E -field component would not have been detected at all by the vertical electric monopole that was used to collect the CCIR noise data. Conversely, the sensitivity of the CCIR test antenna has an azimuthal symmetry, but the corresponding magnetic antenna – the vertical loop – clearly does not; a fact which supports the above assertion that, because the noise fields are not single waves with a unique direction, they cannot be related by Z_0 .

We can also note that, at low frequencies where the ground is a good conductor and the antenna is close to the ground (compared with a skin depth), a direct and a ground-reflected field would tend to cancel. The components suffering the greatest cancellation would be the horizontal electric field and the vertical magnetic field, as we depicted in *Figure 2-10*. This may suggest that the noise received by a VMD is less than the noise received by a VED, (as tabulated by the CCIR).

The above points serve to highlight the fact that the CCIR noise data may not be particularly relevant to us; and also the importance of the magnetic noise data that we wish to collect using the sounding equipment to be described in later chapters. Nevertheless, for the purposes of this initial discussion of atmospheric noise, we are forced – in the absence of any other data – to make the assumption that the fields *are* in the ratio of Z_0 and that the CCIR data *can* be used.

To derive the electric *equivalent noise field spectral density* from a given *atmospheric noise temperature ratio* in a matched r.f. antenna we first consider the noise voltage due to the

defined effect of F_a , that is

$$v = \sqrt{4kT_0 F_a B R_r} \quad (6-7)$$

and then equate this to the noise voltage due to the power flux spectral density \hat{P}_d of the equivalent noise field, which we assume is incident on an antenna of capture cross-section A_c and a gain G ,

$$v = \sqrt{4\hat{P}_d A_c G B R_r} \quad (6-8)$$

This results in the relationship

$$F_a = \frac{\hat{P}_d A_c G}{kT_0} \quad (6-9)$$

which defines F_a in terms of the equivalent field that gives rise to a measured noise voltage. The derivation assumes a matched antenna, but mis-matching would not alter the equivalent noise field and so the concept applies equally to an unmatched quasi-static system. The quasi-static approach is therefore to define the spectral density \hat{H}_n such that the corresponding magnetic field is $\hat{H}_n \sqrt{B}$ and then to equate (6-1a) and (6-1c), to write

$$\begin{aligned} \omega \mu (\hat{H}_n \sqrt{B}) \cdot \eta \cdot \Phi \sqrt{R} &= \sqrt{4kT_0 B F_a R_r} \cdot \eta \\ \Rightarrow F_a &= \frac{(k_0 Z_0 \hat{H}_n)^2}{4kT_0} \left/ \left(\frac{R_r}{R \Phi^2} \right) \right. \end{aligned} \quad (6-10)$$

then, noting that $\hat{H}_n^2 = \hat{P}_d / Z_0$ and that the denominator term is equal to $Z_0 k_0^4 / 6\pi$ (see equation (3-27) defining radiation resistance), we obtain

$$F_a = \frac{\hat{P}_n}{4kT_0} \cdot \frac{6\pi}{k_0^2} \quad (6-11)$$

Since we know that, and that $A_c = \lambda^2 / 4\pi$, $G = 3/2$ for a small magnetic dipole, we can show that (6-9) and (6-11) are equivalent. We can also write (6-11) as

$$20 \log \hat{H}_n = 20 \log f + 10 \log F_a - 390.05 \quad (6-12)$$

which we can use to derive an atmospheric noise temperature ratio from measurements made using a small magnetic dipole antenna. We shall use this equation in the noise examples that follow although, as we discussed above, we are not fully justified in using (6-12) to deduce a magnetic noise field from the CCIR/ITU, which was collected using an *E*-field antenna.

It is, however, valid to derive the electric equivalent noise field spectral density \hat{E}_n by adding $20 \log(Z_0) = 51.52$ to (6-12), provided that we then use the equation to derive the atmospheric noise voltage resulting from a small *electric* dipole. If we were measuring the

noise in a self-resonant half-wave dipole with gain⁹ $G = 1.64$ then there would be a further factor of $10 \log(1.64 / 1.50)$ to subtract from the right-hand side of (6-12), which would result in

$$\Rightarrow \hat{E}_n \Big|_{dB[\mu V/m]} = 20 \log f \Big|_{MHz} + 10 \log F_a - 98.92 \quad (6-13)$$

which is the formula quoted in ITU report 372.

6.2.2 Typical Values for Atmospheric Noise

Generally, there can be significant seasonal, diurnal or geographic variations in the atmospheric noise level and, at an atmospherically-quiet receiving site, man-made noise may predominate. The noise does not obey a simple relationship with frequency and the variation of the noise with location and time is smaller below 10kHz.

The ITU report contains graphs of world-wide atmospheric noise at 1MHz, together with conversion factors for use at other frequencies. In some circumstances the atmospheric noise can be lower than man-made noise, in which situation the quoted figures are estimates. *Figure 6-1* and *Figure 6-2*, below, are copies of figures from ITU report P.372-7, and show the generalised world-wide data. For our purposes, we can make some crude approximations as shown in table *Table 6-1*, below.

6.2.3 Typical Values for Man-Made Noise

From 30kHz to 250MHz, according to the CCIR/ITU data, the noise temperature of man-made noise shows a uniform logarithmic variation with frequency. Table 1 and Figure 10 of the ITU report can be summarised by saying that in rural areas F_a is typically 95dB at 100kHz, falling smoothly at around 28dB/decade. However, these measurements were made in the 1970s and may no longer be appropriate. A further summary of the ITU report is in *Table 6-2* below, which is included for comparison with further atmospheric noise data *Table 6-3* below.

Key to Figure 6-2 on page 136

- A – Atmospheric noise, value exceeded 0.5% of time
- B – Atmospheric noise, value exceeded 99.5% of time
- C – man-made noise, quiet receiving site
- D – galactic noise
- E – median business area man-made noise
- heavy line indicates minimum noise level expected.

9 See for example, (Ramo, Whinnery and Van Duzer, 1984) §12.6

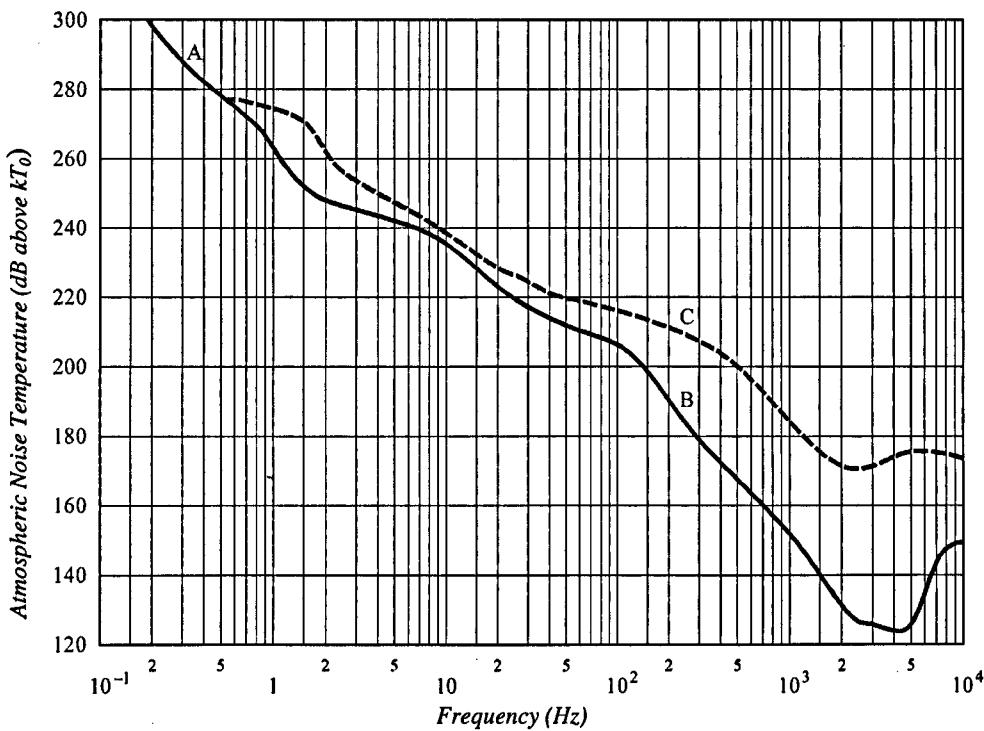


Figure 6-1 – Expected value of world-wide atmospheric noise below 10kHz.

Y-axis: Noise temperature (dB above kT_0). X-axis: Frequency (Hz).

Key: A – Micro-pulsations, B – minimum expected value, C – maximum expected value.

Reproduced from ITU report P.372-7, figure 1.

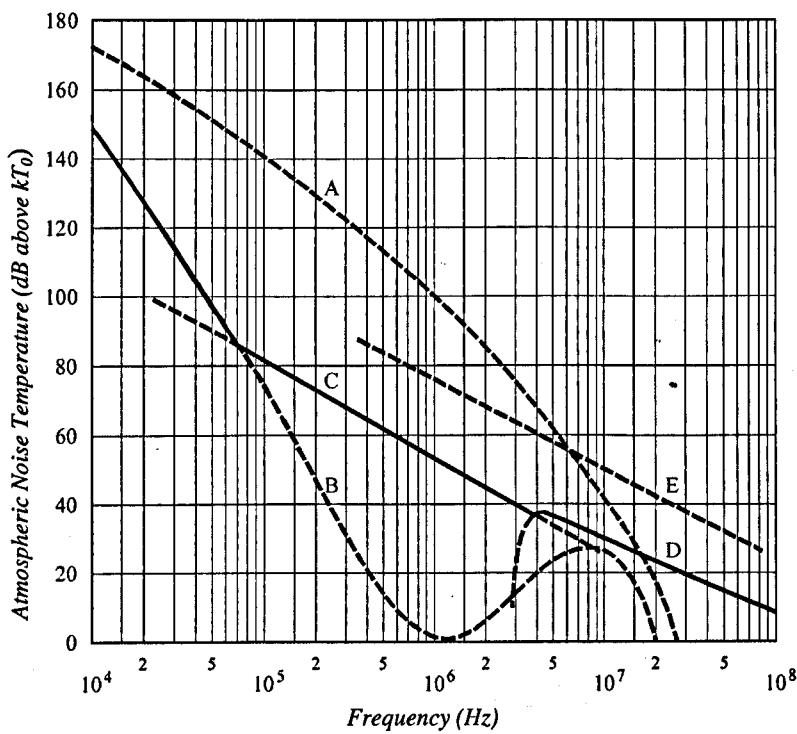


Figure 6-2 : Expected noise temperature from 10kHz to 10MHz.
Y-axis: Noise Temperature (dB above kT_0). X-axis: Frequency (Hz).

See foot of previous page for Key.

Reproduced from ITU report P.372-7, figure 2.

Type of Noise	Value of F_a	Change in F_a (dB/decade)	
		< 10kHz	10kHz – 1MHz
atmospheric (maximum)	172dB at 10kHz	-23	-36
atmospheric (minimum)	148dB at 10kHz	-31	-74
man-made	124dB at 10kHz	-	-28

Table 6-1 : Likely range of Atmospheric Noise Temperature Ratio, F_a .
ITU data taken from Figure 6-1 and Figure 6-2 above.

Environment	Median value of man-made noise temperature ratio F_a , in dB above kT , at...			
	20kHz	150kHz	200kHz	1MHz
Business	124	100	96	77
Residential	120	95	92	73
Rural	114	90	87	67
Quiet Rural	102	77	74	54
(Galactic)	91	71	68	52

Table 6-2 : Man-made noise level in different environments.
Data taken from ITU report P.372-8 ((ITU, 2001))

6.2.4 Values of Noise in the UK

ITU report P.372-7 includes 24 contour maps showing the world-wide noise temperature ratio, at 1MHz, for different times of day (4 hour periods) and different seasons¹⁰. For each map there is an accompanying conversion chart that allows the noise at other frequencies to be ascertained; and a chart that allows the standard deviation and upper/lower decile values to be determined. Interpretation of this data requires care, because it is presented for ‘local’ time and ‘local’ season (the latter explaining the discontinuity of the data at the equator). In his book *VLF Radio Engineering*, (Watt, 1967) discusses atmospheric noise in detail, with particular reference to LF noise, and he reproduces the 24 CCIR graphs. But Watt presents the data with respect to GMT instead of local time, and to month instead of season (e.g. March–May instead of Spring); and he presents contours for electric noise field at 10kHz (dB above 1V/m/Hz) rather than as a noise temperature ratio at 1MHz. Thus, on initial inspection, Watt’s data bears little or no resemblance to the earlier CCIR data.

The data in *Table 6-3* on page 138 was interpolated from the graphs and charts in CCIR report 332 by (Rabson, 1993)¹¹ and shows the mean value of the atmospheric noise temperature

¹⁰ ITU-R P.372-7 has been compiled by electronically scanning the graphs from CCIR 322, and much of the detail has not been reproduced. Therefore, for serious study, we suggest consulting a library copy of the earlier CCIR 322 report.

¹¹ Reproduced by permission, with corrections as noted in Table caption.

ratio, F_a in southern England. We can observe that, although there is a wide seasonal and diurnal variation at 1MHz, this is much reduced at 20kHz. It is also apparent that, for much of the daylight hours, this CCIR estimate of atmospheric noise (i.e. lightning strikes and electrical storms) is actually lower than the likely value for man-made noise.

Season	Local Time	Median value of expected atmospheric noise temperature ratio F_a , in dB above kT , at...			
		20kHz	150kHz	200kHz	1MHz
Winter	00-04	† 145	106	100	70
	04-08	143	98	92	63
	08-12	134	75	66	26
	12-16	136	79	71	34
	16-20	142	95	90	60
	20-24	143	104	96	68
Spring	00-04	145	104	97	67
	04-08	140	88	80	45
	08-12	138	80	71	30
	12-16	140	86	78	37
	16-20	142	92	85	52
	20-24	145	104	98	68
Summer	00-04	147	106	97	67
	04-08	144	95	86	48
	08-12	148	90	80	34
	12-16	145	90	82	40
	16-20	145	96	† 90	54
	20-24	146	105	97	67
Autumn	00-04	147	107	100	70
	04-08	144	97	90	56
	08-12	140	86	77	33
	12-16	142	89	82	40
	16-20	144	100	93	60
	20-24	146	108	102	72

Table 6-3 : Seasonal atmospheric noise temperature in southern England.

Data from CCIR report 322, summarised by (Rabson, 1993);
values marked † have been corrected by Gibson.

Shaded values indicate that the expected noise level due to atmospheric phenomena is lower than the published values for man-made noise at a rural location.

6.2.5 Antenna Size and Atmospheric Noise Term, F_k .

In §3.4.1.1 we found it convenient to define an ‘atmospheric noise term’, F_k as

$$F_k = \sqrt{F_a \frac{Z_0 k_0^4}{6\pi}} \quad (6-14)$$

$$\Rightarrow 20 \log F_k = 10 \log F_a + 40 \log f - 294.147$$

F_k is a useful parameter because it has the reciprocal dimensions to specific aperture and allows us easily to determine the necessary size of an antenna. In §7.2.3 and §7.2.4 we summarised the likely maximum and minimum values for external noise which are as shown in *Table 6-4*, in terms of F_k , and plotted in *Figure 6-3*. Notice that, although the atmospheric noise temperature reduces monotonically with increasing frequency, this is not so for F_k . The minimum value of F_k , in this figure, is -4.5dB at 35kHz . This suggests that the maximum useful size of the receiver antenna under all conditions is $3.4\text{m}^2/\sqrt{\Omega}$, although a larger antenna can be used in conditions of lower noise.

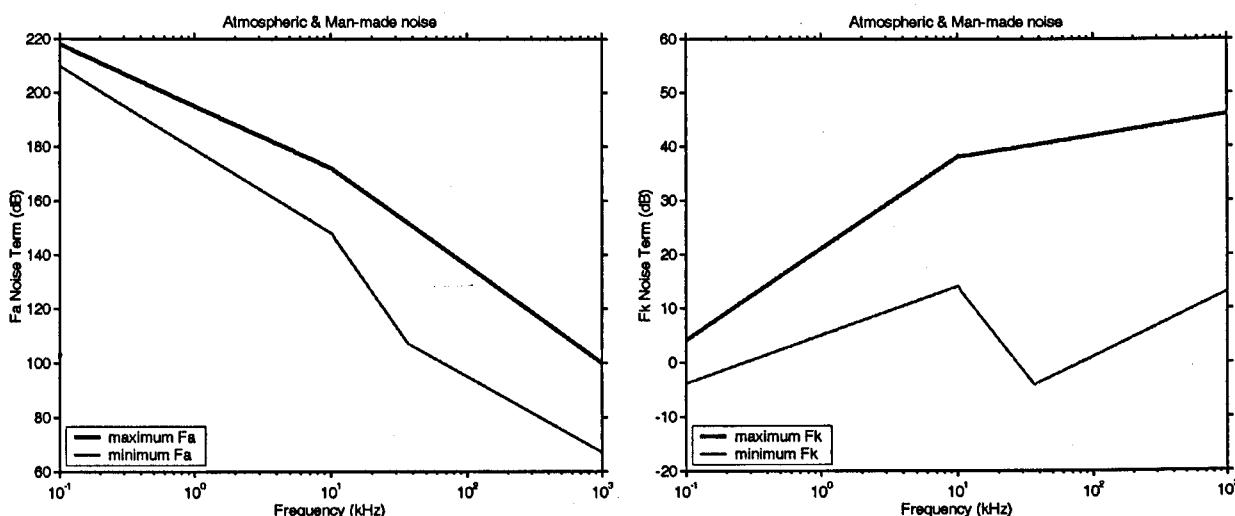


Figure 6-3 : Likely range of noise terms F_a and F_k .
Data from Table 6-4 and earlier discussion.
Note that although F_a falls monotonically with frequency, F_k can actually rise with frequency.

Type of Noise	Value of F_k	Change in F_k (dB/decade)	
		< 10kHz	10kHz – 1MHz
atmospheric (maximum)	+38dB at 10kHz	+17	+4
atmospheric (minimum)	+14dB at 10kHz	+9	-34
man-made	-11dB at 10kHz	-	+12

Table 6-4 : Likely range of noise term F_k .
(Compare with Table 6-1)

6.2.5.1 Example: high noise conditions

The noise field at 150kHz , due to a noise temperature ratio of 100dB (see *Table 6-3*), is found from (6-12) to be $-186\text{dB[A/m/\sqrt{Hz}]}$. Using (6-14) we can derive $F_k \approx 4.4\sqrt{\Omega}/\text{m}^2$. For the atmospheric noise term to be the dominating term in the expression for noise factor, we must

have $F_k \Phi > 2$, so we can immediately state that our receiver antenna need be no larger than $0.45\text{m}^2/\sqrt{\Omega}$. This is an extremely small antenna; for example, it could be

- a) an air-cored loop, diameter 300mm, with 3 turns of 7/0.2mm equipment wire
- b) a ferrite-cored solenoid, 8mmØ × 200mm, with 110 turns of 0.1mmØ copper wire.
(with the effective relative permeability being shape-factor-limited to 200).

Given that an increase in antenna size will not improve the SNR, we can deduce that, under these conditions, we must use a more powerful transmitter to overcome the atmospheric noise.

6.2.5.2 Example: low noise conditions

In quiet conditions, we might expect the noise temperature ratio to drop to 75dB at 150kHz – a reduction of 25dB from the above example. If F_a falls by 25dB, F_k will fall by about 18 times and so the maximum useful size of our receiver antenna rises by 18 times to $8\text{m}^2/\sqrt{\Omega}$, which is more in line with the typical antennas discussed in chapter 5. This reduction in atmospheric noise of 25dB is similar to the reduction we would observe if we were to operate at around 2.8 skin depths where, as we discussed above, we can make use of a larger antenna to take advantage of the higher SNR.

6.2.6 Noise Performance: Example

As a brief example we will consider a communications system at 20kHz.

6.2.6.1 Typical Antenna

We will consider the antenna described in *Table 4-4* as ‘Loop 2’ and assume this is used horizontally (i.e. a vertical magnetic dipole, VMD) with the transmitter and receiver arranged co-axially. The loop consists of 34 turns of 7/0.2mm copper wire wound on a diameter of 1m, and we can also describe it as having a specific aperture of $9.2\text{m}^2/\sqrt{\Omega}$ and a resistance of 8.4Ω . Its signal bandwidth is given in the table as 185Hz and so the Q -factor at 20kHz will be 108^{12} .

6.2.6.2 Received Noise Voltage

We will assume that the atmospheric noise temperature ratio is 148dB (see *Table 6-3*) although we should note, from the earlier discussion, that this figure was derived for a vertical electric monopole and it is unlikely to apply to our VMD antenna. We will also assume, for the purposes of this example, that the electric and magnetic noise fields are in the ratio Z_0 . Again, as discussed earlier, this will not generally be the case. The magnetic noise field spectral density is found from (6-12) to be -156dB , or $15.8\text{nA/m}/\sqrt{\text{Hz}}$. The electric noise field is found

¹² These calculations, and those presented below, were obtained using a spreadsheet which is included on the bound-in CD-ROM – see appendix A9.

(noting the caveats above) by adding 51.5dB to the value of the magnetic noise field, to give – 104dB[V/m/Hz]. From (6-14) we can also derive $F_k = 25.9\text{dB}$ or $19.7\sqrt{\Omega/\text{m}^2}$

It is straightforward to show that the received voltage that results from a magnetic noise field spectral density, \hat{H}_n , when detected by a parallel-tuned resonant antenna, is given by

$$v_n|_{\substack{\text{tuned} \\ \text{antenna}}} \approx \omega \mu \hat{H}_n \cdot NA \cdot \sqrt{\frac{1}{2}\pi B_s} \cdot Q \quad (6-15)$$

where the antenna has N turns and area A ; Q -factor Q and signal bandwidth B_s (noting that the noise-equivalent bandwidth is $\frac{1}{2}\pi B_s$). The approximation in the above equation arises because we have assumed that the frequency is constant, instead of integrating over the bandwidth. Also, it has been the practice, in this thesis, to describe the antenna in terms of its specific aperture Φ , instead of using N and A , which we can do by using (4-7).

Similarly, the electric noise field, incident on a whip antenna, will result in a voltage of

$$v_n|_{\substack{\text{tuned} \\ \text{antenna}}} \approx \hat{E}_n \cdot h_e \cdot \sqrt{\frac{1}{2}\pi B_s} \cdot Q \quad (6-16)$$

where h_e is the effective height of the antenna. The electric field would usually be rejected by a combination of an electrostatic screen and a differential amplifier, but we could suppose that inadequate shielding and balancing would result in the loop acting as an E -field antenna with an effective height of, say, 10mm, for the purposes of this example.

The thermal noise voltage of the antenna is simply

$$v_n|_{\substack{\text{tuned} \\ \text{antenna}}} = \sqrt{4kTR} \cdot \sqrt{\frac{1}{2}\pi B_s} \cdot Q \quad (6-17)$$

6.2.6.3 Received Signal Voltage

At 20kHz, with a ground conductivity of 1.25mS/m, the skin depth will be 100m. Therefore operating at 250m ($T = 2.5$), and using the factor of $\exp(-T) \cdot \sqrt{T^2 + (1+T)^2}$ from equation 2-25 we can state that the signal would be reduced by 9.0dB more than that predicted by the inverse cube law alone. If a transmitter were to use the same antenna as the receiver, above, then it would produce a magnetic moment of 75Am^2 for a power dissipation of 66W, resulting in a coaxial field of around 270nA/m at a distance of 250m; and a signal voltage of $123\mu\text{V}$ r.m.s. In passing we can observe that the skin depth attenuation requires us to use 9dB more power (i.e. 8 times) that would otherwise be the case.

Summarising this data, in *Table 6-5*, we can see that there is a significant noise voltage due to the electric component of the noise field; thus demonstrating the necessity of adequately screening and electrically balancing the induction loop. The significance of the noise voltage

can only be appreciated when we consider the overall signal/noise ratio (SNR) of the receiver which depends, of course, on the strength of the transmitted signal.

	Noise temp. ratio	Noise field density at 20kHz	Induced voltage	...with antenna of...
Magnetic field	148dB	-156dB, 15.8nA/m/ $\sqrt{\text{Hz}}$	123 μV	34 turns, 1mØ
Electric field		-104dB, 5.96 $\mu\text{A}/\text{m}/\sqrt{\text{Hz}}$	110 μV	height 10mm
Thermal noise	-	-	0.67 μV	34 turns, 1mØ
Signal		Moment 75Am^2 , @ 250m, with $T=2.5$; results in field of $270\text{nA}/\text{m}$	123 μV	34 turns, 1mØ

Table 6-5 : Example: receiver antenna voltage due to signal and sources of noise.
Figures are for the antenna described in §6.2.6.

6.2.6.4 Adequate Size of Antenna

The receiver antenna in this example has $\Phi = 9.2\text{m}^2/\sqrt{\Omega}$, and the atmospheric noise can be represented as $F_k \approx 20\sqrt{\Omega}/\text{m}^2$, so the antenna easily meets the criterion $-F_k \Phi > 2$ – for being ‘large’ and – as we can see from the table – its SNR is indeed dominated by the atmospheric noise. At the quoted range of 250m the SNR is 0dB and we can verify that this agrees with the formula for the SNR of a ‘large’ antenna, given in §3.4.1.1 and (6-18) below. Additionally, the signal to thermal noise ratio is 45dB and this, too, can be found using the formulas in §3.4.1.1.

6.2.7 Up-link v. Down-link Operation

There is an obvious difference between up-link and down-link operation, which is that the up-link system presents the full value of atmospheric noise to the receiver, whereas the down-link operation results in the atmospheric noise being attenuated as it propagates down through the ground at 8.7dB per skin depth. We saw, in chapter 2, that the optimum operating distance was about 2.8 skin depths, which does not depend on the level of atmospheric noise. Down-link communications at the optimum distance can therefore be expected to have a 25dB greater SNR than up-link similar communication. In this situation, is clearly advantageous to operate at shorter distances and we must remind ourselves that ‘optimum distance’ is a misleading description – the situation is that *given* a particular distance, there is an *optimum frequency* which maximises the SNR.

6.2.7.1 Range: Up-link Operation

The calculation in the previous section derived an SNR of 0dB. This is a convenient value to use when describing the sensitivity of an antenna, but it is too low for operational use. Let us

assume an operating distance of 125m. The signal strength will be 8 times (18dB) greater due to cube-law attenuation, and only 2.6dB instead of 9dB due to skin depth attenuation, resulting in an SNR, for the up-link operation of 24.4dB. This is shown, diagrammatically, in *Figure 6-4* below.

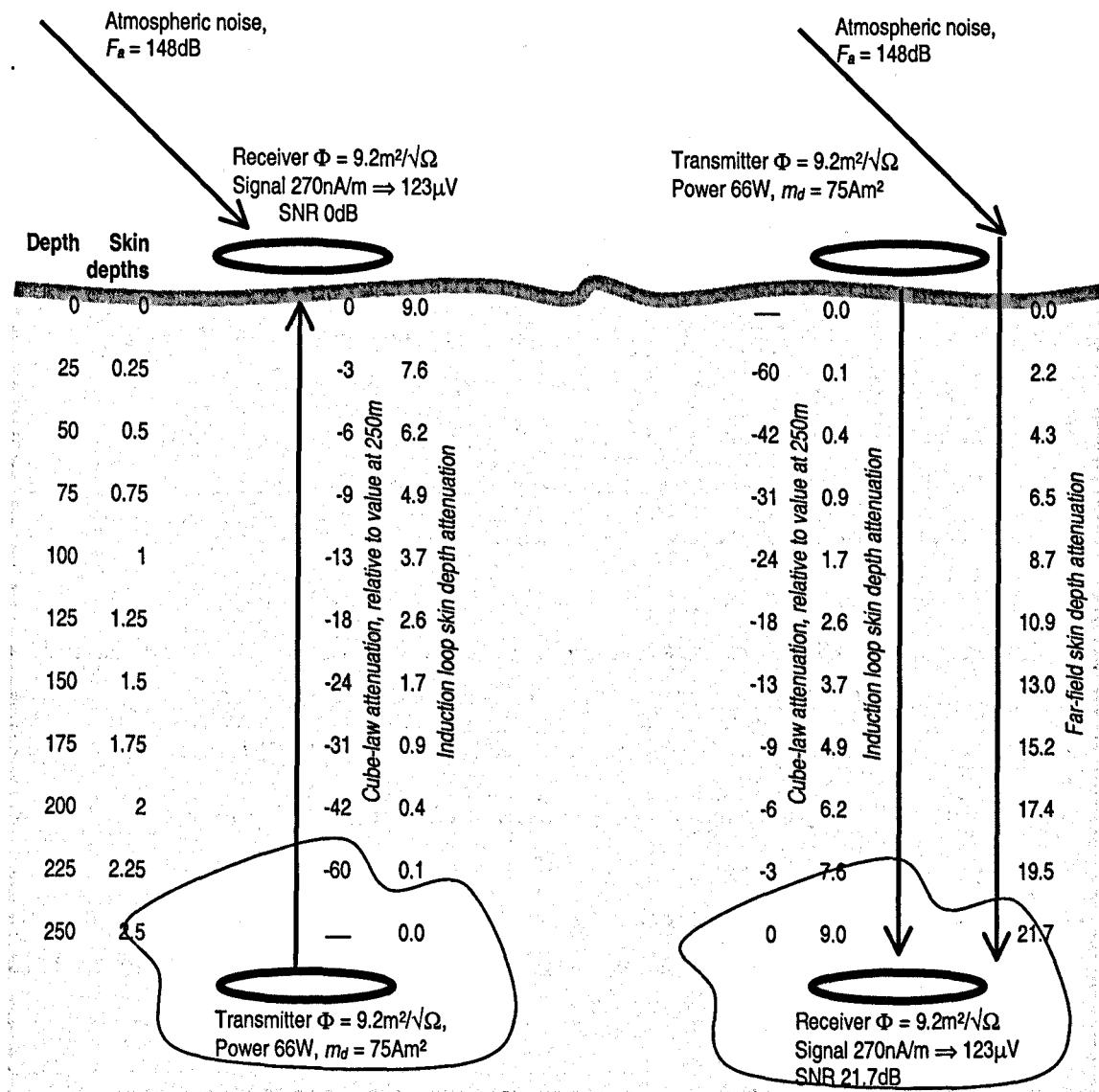


Figure 6-4 : A comparison of up-link and down-link SNR.
With the data given in the text, the up-link SNR is 0dB at 250m, whereas the down-link SNR is 21.7dB. At 125m the up-link improves to 24.4dB compared with 35.3dB for the down-link.

6.2.7.2 Range: Down-link Operation

For the down-link operation, we can assume that the atmospheric magnetic noise is attenuated at 8.7dB per skin depth so the SNR at the subsurface receiver is increased by nearly 10.9dB if we are operating at 125m ($T = 1.25$). However, this improvement is not straightforward to interpret. Firstly, the high refractive index of the rock means that if the noise field is incident on a horizontal uniform medium, the wave will travel almost vertically

downwards. Since the **E** and **H** vectors are perpendicular to the direction of propagation, it follows that the vertical component of the noise field is almost zero, so a horizontal loop antenna should detect very little atmospheric noise. This particular situation may not occur in reality so, for his example, we will assume the figure of 10.9dB mentioned above.

10.9dB less atmospheric noise does not necessarily mean an increase of 10.9dB in SNR, because of the contribution from the antenna and amplifier. However, with F_k reduced by almost 11dB to $5.6\sqrt{\Omega/m^2}$, we can see that the $F_k \Phi$ product is about 52 and so, if we have correctly designed the receiver then the atmospheric noise will still be the most significant contribution to SNR.

We can therefore deduce that we have an SNR of 35.3dB at a range of 125m. Alternatively, if we wished to maintain the SNR at 24.4dB, then the transmitter power could be reduced by 10.9dB, to around 5W. This is a significant amount, corresponding to an increase in battery life of over 12 times. Instead of reducing the transmitter power, the sub-surface receiver antenna could be 10.9dB or 3.5 times smaller (i.e. have a specific aperture 3.5 times lower) – provided that the $F_k \Phi$ product is still sufficient, which it would be in this example.

A third way to decrease the SNR to the 24.4dB of the up-link example is to increase the range. However, this is not as trivial to calculate. We recall that, at 250m, the signal attenuation was 24dB greater (§6.2.7.1) than at 125m. The atmospheric noise will be attenuated by a further 11dB and so the SNR will fall by the difference, resulting in an SNR of 21.7dB.

Now we can note a significant difference in the down-link range, compared with the up-link because, at 250m, we noted the up-link had an SNR of 0dB. We can also note that, with the additional attenuation of the atmospheric noise, the $F_k \Phi$ product will be reduced to around 15, indicating that the atmospheric noise is still the dominant contribution. In fact, it is not until a depth of around 450m, or 4.5 skin depths, that $F_k \Phi$ drops to a level at which the thermal noise of the antenna is significant. At 450m the down-link SNR is still a useable 10.7dB, whereas the up-link SNR is -28.4dB.

6.2.8 Optimum Frequency in the Presence of Noise

We have seen that, with the atmospheric noise predominant, the SNR is independent of the size of the receiver antenna and can be expressed, from (6-2) and (6-5), as

$$\frac{S}{N} = \frac{\omega \mu H_s}{\sqrt{4kTB}} / F_k \quad (6-18)$$

There are two interesting ‘special cases’ concerning the term F_k . We have seen (6-14) that the square of F_k is proportional to F_a multiplied by the fourth power of frequency, from which we can deduce that

- i) if the noise temperature ratio F_a falls at $-40\text{dB}/\text{decade}$ then the F_k term will be independent of frequency; and
- ii) if F_a falls at $-20\text{dB}/\text{decade}$ then F_k will be proportional to frequency, and so the signal/noise ratio will be independent of frequency.

These observations are important because they suggest that the large increase in F_a at low frequencies might not result in a decrease in SNR. We will now consider this in more detail.

(6-18) described the SNR in terms of the magnetic field strength of the received signal H_s but we can also write it in terms of the transmitted signal by modifying (2-25) to give

$$\left. \frac{S}{N} \right|_{\theta=0} = \frac{m_d}{2\pi\sigma} \frac{1}{\sqrt{4kTB}} \frac{1}{r^5} T^2 \cdot 2e^{-T} \sqrt{T^2 + (1+T)^2} / F_k \quad (6-19)$$

If F_a varies as ω^{-n} (i.e. it falls at $10n\text{ dB/decade}$) then we can replace $1/F_k$ by the term T^{n-4} (making use of the definition of skin depth as a function of ω) to obtain

$$\left. \frac{S}{N} \right|_{\theta=0} \propto T^{n-2} e^{-T} \sqrt{T^2 + (1+T)^2} \quad (6-20)$$

This demonstrates an optimum frequency window, in a similar fashion to the earlier analysis in §2.2.4, and is plotted in *Figure 6-5*, below. We can interpret the figure by first considering the bold line, and assuming we are operating the optimum frequency point, $T \approx 2.83$. If the atmospheric noise at this frequency remains unaltered, but varies either side of this point at a rate of -30dB/decade then, considering the line marked ' -30dB ', we can see that there is a slight shift in the optimum frequency but that the Q -factor of the curves does not vary by much.

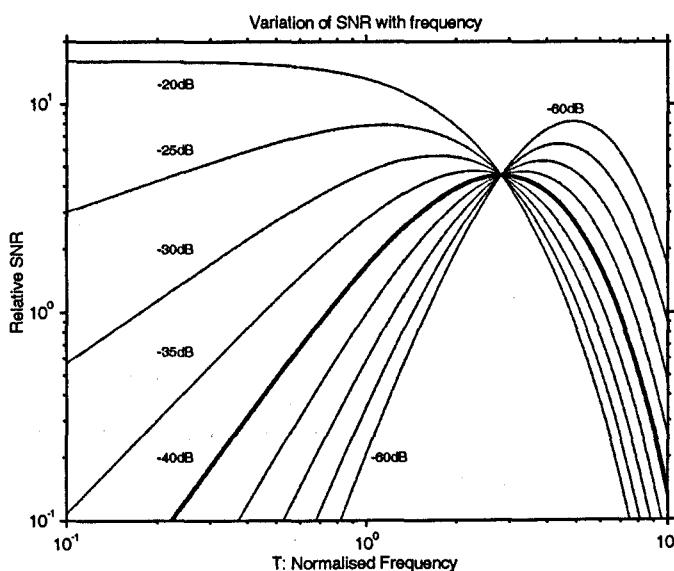


Figure 6-5 : Variation of SNR with frequency and atmospheric noise temperature profile. The labels indicate the variation of F_a per decade of frequency. If this falls at 40dB/decade the optimum frequency window is unaffected by the noise and so the curve matches that Figure 2-9. SNR is relative to an arbitrary datum.

Should we wish to calculate the optimum frequency, it can be found by differentiating (6-20) and setting to zero, to give

$$\frac{2}{n-2}T^3 - 2T^2 - 2T - 1 = 0 \quad (6-21)$$

which, for $n = 4$, becomes identical to (2-26) as we would expect from the above discussion.

The likely variation of F_k with frequency was shown in *Figure 6-3*. This noise data can be applied to (6-19) and used to produce the graph of *Figure 6-6*, below, which shows the signal/noise ratio a function of frequency, for a range of depths. The situation being modelled is that of up-link communication using co-axial horizontal loop antennas (i.e. vertical magnetic dipoles), with a ground conductivity of 1mS/m. The simple model of propagation has been used, rather than Wait's full treatment, because we showed, in chapter 2, that there was little difference in the result for this configuration of antennas.

The reference point for the SNR in *Figure 6-6* is arbitrary, and has been chosen for a transmitter of magnetic moment 30Am^2 and a receiver noise-equivalent bandwidth of 1000Hz. The left figure shows the 'no noise' curve for comparison with *Figure 2-18*. The data for this curve is essentially that for the maximum noise, but varying at -40dB/decade which, as we saw above, results in a value for F_k that does not vary with frequency.

We can draw some important conclusions from this data. Firstly, if we are indeed free to choose to operate at the supposed 'optimum frequency' then the variation of atmospheric noise temperature with frequency is unlikely to dictate a severe shift in the optimum frequency, because of the flat top to the curves. A related point is that if we are considering wideband communications, the bandwidth of the optimum frequency window response is not unduly affected by noise – that is, although the noise temperature may be higher at lower frequencies, the received noise voltage shows a more uniform distribution with frequency.

From *Figure 6-6* we can see that, for the parameters used, the optimum frequency for a depth of 100m is 148kHz under maximum (or no) noise conditions¹³. But if the noise has the profile associated with the minimum figures then the optimum frequency is 58kHz. However, in this situation, the SNR at 148kHz is still higher than it would be with maximum noise. Similar observations can be made throughout the spectrum and we can conclude that if we choose the optimum frequency on the basis of signal propagation alone, then the effect of atmospheric noise is unlikely to dictate a change to that frequency.

¹³ This is using the approximate model. Wait's model gives figures of 162kHz, with 68kHz for the minimum noise situation. Because of the low Q -factor of the optimum frequency window, these differences are not significant.

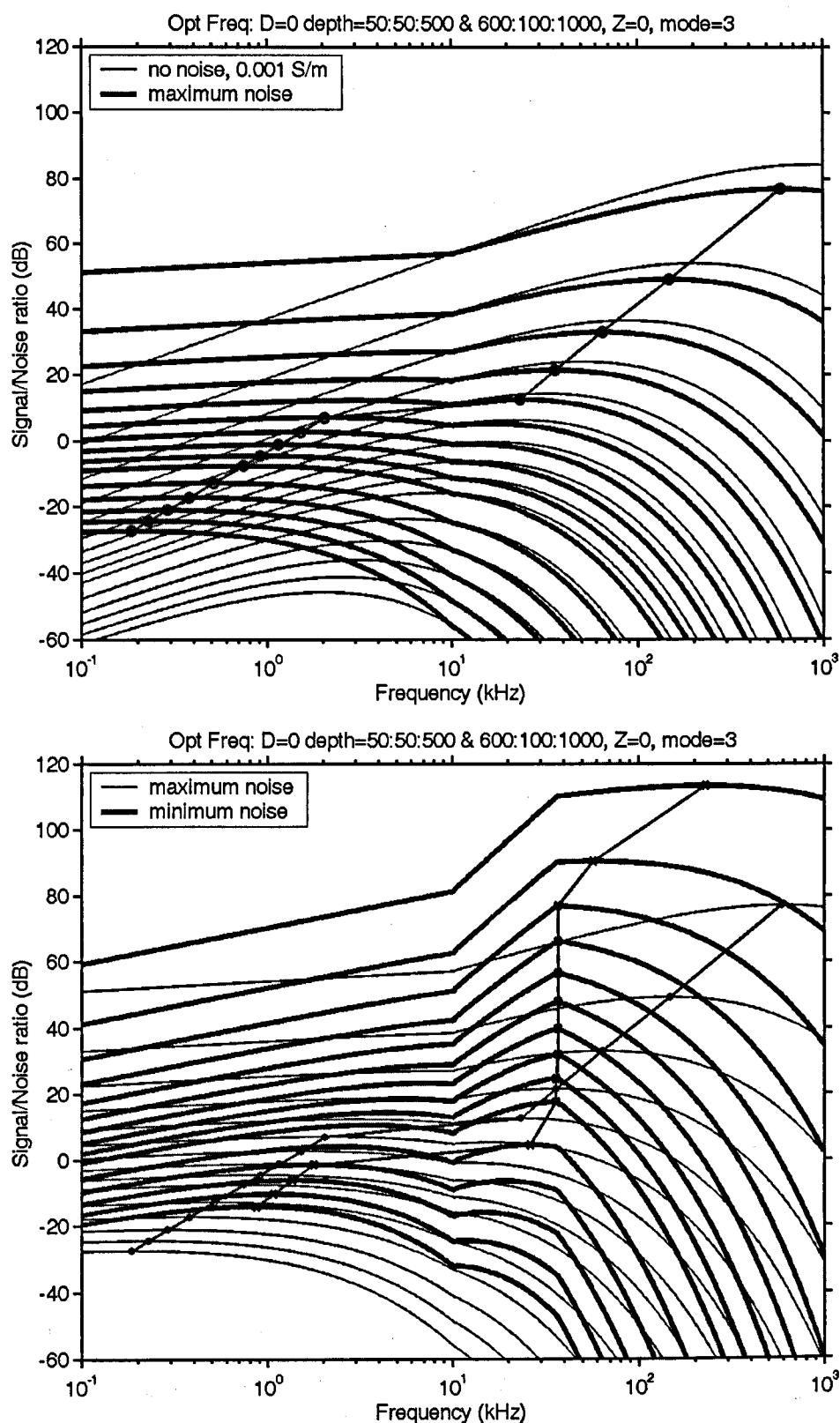


Figure 6-6 : Variation of SNR with frequency for different depths.

Co-axial antennas, up-link communications.

Depth: (top to bottom curves) 50m–500m step 50m, 600–1000m step 100m

Top: maximum noise condition; also shows 'no noise' curve for comparison (see text).

Bottom: minimum noise condition; also shows minimum noise for comparison.

Locus of optimum frequency condition shown by —●—●—

SNR is relative to an arbitrary datum – see text.

6.3 Noise Factor of a Receiver Antenna

We have already seen, in §6.1.2.2, how the amplifier's input-referred noise voltage and input noise current affect the noise factor of the receiver. However, the earlier derivation assumed that the antenna could be represented as a resistance. In practice, this may not be so – for example, if the antenna were an untuned inductor then it would present a highly reactive source impedance, which would accentuate the noise contribution due to the amplifier's input noise current. We will now analyse a number of antenna configurations; the results will be summarised in *Table 6-6* on page 152. The SNR is, in all cases, described by (6-2) and (6-5), with the noise factor given by

$$F_n = 1 + (F_k \Phi)^2 + \frac{\hat{P}_n}{4kT_0} A_n \quad (6-22)$$

where A_n represents the noise contribution of the amplifier, and contains the noise matching terms. The value of A_n is given in the table for various specific conditions.

We will consider four basic antenna tuning configurations, as shown in *Figure 6-7* below.

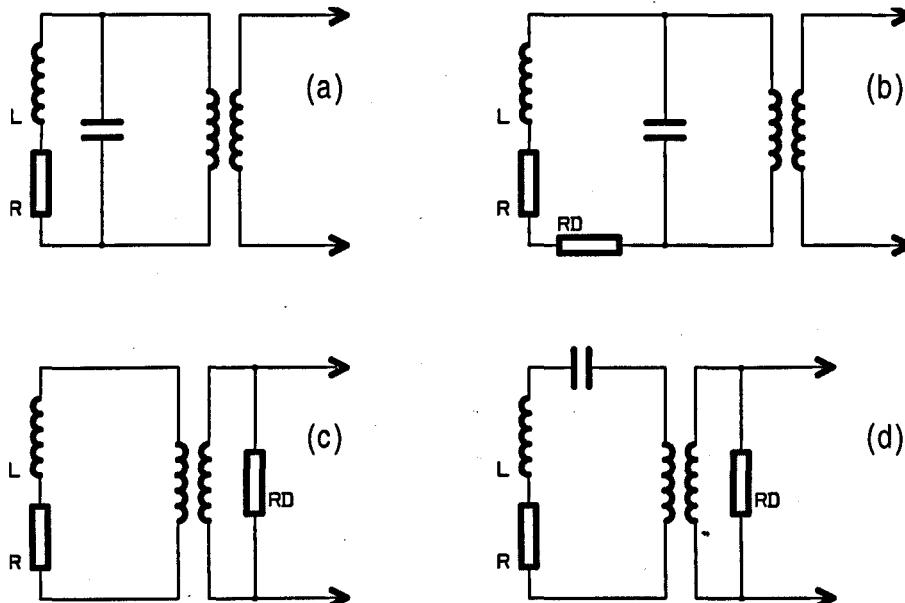


Figure 6-7 : Tuned and untuned antenna configurations.
The antenna consists of elements L and R and a noise-matching-transformer.
Circuit a) parallel-tuned, b) parallel-tuned with damping resistor R_d ,
c) untuned with damping resistor, d) series-tuned with damping resistor.

Noting that the tuned antenna has impedance Z , which consists of elements R (the loop resistance), L and C , we will make the convenient definition that

$$Q_0 = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 C R} \quad (6-23)$$

where ω_0 is the resonant frequency that the *LCR* circuit would have in isolation from any other circuit elements (such as a damping resistor, which we may introduce), and Q_0 is used to indicate that the *Q*-factor is defined at this frequency. It will be convenient to use some shorthand notation, writing

$$\bar{\Omega} = \Omega - \frac{1}{\Omega}, \quad \text{with} \quad \Omega = \frac{\omega}{\omega_0} \quad (6-24)$$

6.3.1 Parallel-tuned Antenna

We will begin by analysing the noise sources in a tuned antenna system, *Figure 6-7a*. This is easier to analyse than the untuned case, because we can assume that the antenna is entirely resistive although this is, of course, only strictly the case at the resonant frequency.

6.3.1.1 Noise Factor at Resonance

Provided that the *Q*-factor of the antenna is high, the analysis follows that given in §6.1.2.2 with the modification that the signals presented by the antenna are Q_0 times greater. The *noise-matching* condition becomes

$$m_d = R\eta^2 Q_0^2 \quad (6-25)$$

and we can achieve this by constructing a matching transformer with the required turns ratio, although this device may, of course, itself introduce noise and distortion, due to non-linearity in the magnetic material.

6.3.1.2 Noise Factor Away From Resonance

The detail of this calculation is shown in appendix A4.2 and we can write the off-resonance noise factor, as

$$F_n = 1 + (F_k \Phi)^2 + \frac{P_n}{4kT} \left\{ \frac{R_n}{\eta^2 R Q_0^2} \sqrt{1 + (Q_0 \bar{\Omega})^2} \right\} \sqrt{1 + (Q_0 \bar{\Omega})^2} \quad (6-26)$$

with $Q_0 \bar{\Omega} \gg 1$, that is, $\omega L / R \gg 1$

where $\bar{\Omega}$ is as defined in equation (6-24).

6.3.1.3 Noise Factor of Damped Antenna

We can also consider the effect of adding a damping resistor R_d in series with the inductor, *Figure 6-7b*. This will broaden the bandwidth by θ times, where

$$\theta = 1 + \frac{R_d}{R} \quad (6-27)$$

Since the only purpose of damping the antenna is to increase its bandwidth, we must consider the off-resonance condition. We will make the assumption that we apply only a limited amount of damping, and the noise factor (appendix A4.2) becomes

$$F_n = \theta + (F_k \Phi)^2 + \frac{P_n}{4kT} \left\{ \frac{R_n}{R \cdot \eta^2 Q_0^2} \sqrt{\theta^2 + (Q_0 \bar{\Omega})^2} \right\} \sqrt{\theta^2 + (Q_0 \bar{\Omega})^2}$$

with limited damping of $\frac{\theta}{Q_0 \bar{\Omega}} \ll 1 \Rightarrow Q_0 \gg \theta$

(6-28)

6.3.2 Series-tuned Antenna

Before considering a completely untuned antenna (*Figure 6-7c*), it is convenient to look at the case of a series-tuned inductor that is damped by a resistor R_d across the amplifier input, *Figure 6-7d*, from which the untuned antenna can be stated as a special case. Again, the detail of these calculations is left to appendix A4.3.

We note that a high value of R_d will have the effect of damping the tuned circuit, such that the Q -factor will reduce by an amount given by

$$\theta = 1 + \frac{R_d}{\eta^2 R} \quad (6-29)$$

we also assume that $\theta \gg 1$, which results in the noise-matching term of

$$A_n = \frac{R_n}{R \cdot \eta^2} \left(1 + (Q_0 \bar{\Omega} / \theta)^2 \right) + \frac{R \cdot \eta^2}{R_n} \left(1 + (Q_0 \bar{\Omega})^2 \right) \quad (6-30)$$

leading to

$$F_n = 1 + (F_k \Phi)^2 + \frac{P_n}{4kT_0} \left\{ \frac{R_n}{\eta^2 R} \sqrt{\frac{1 + (Q_0 \bar{\Omega} / \theta)^2}{1 + (Q_0 \bar{\Omega})^2}} \right\} \sqrt{1 + (Q_0 \bar{\Omega} / \theta)^2} \sqrt{1 + (Q_0 \bar{\Omega})^2}$$

with $\theta \gg 1$

(6-31)

At the centre frequency of this tuned and damped antenna we will have $\bar{\Omega} = 0$ so the matching condition is quite simple. At the edges of the band (-3dB points) we will have $Q_0 \bar{\Omega} / \theta = 1$. Formulas summarising these conditions appear in *Table 6-6* below.

6.3.3 Untuned Antenna

For an untuned antenna we simply replace $\bar{\Omega}$ by Ω in (6-31). At the -3dB point (the antenna is now a low-pass filter) we again have $Q_0 \Omega / \theta = 1$. However, it is more convenient to write $Q_0 \Omega$ in terms of the reactance of the antenna, as X / R and we can note that, for the situation where the operating frequency is low (i.e. $Q_0 \Omega / \theta \ll 1$), we have

$$F_n = 1 + (F_k \Phi)^2 + \frac{\hat{P}_n}{4kT_0} \left\{ \begin{array}{l} \left\| \frac{R_n}{\eta^2 X} \right\| \frac{X}{R} \\ \text{with } \theta \gg \frac{X}{R} \gg 1 \end{array} \right\} \quad (6-32)$$

We have assumed that the damping resistor, R_d , has a high value but, for completeness, we should also consider the situation where it has a very low value. In this case, the signal (and atmospheric noise) attenuation could be severe, but the amplifier's noise contribution will also be low.

6.3.4 Summary and Observations

The formulas presented above are summarised in *Table 6-6* on page 152. We will comment on some specific antenna configurations, as follows.

6.3.4.1 Parallel-Tuned Antenna

Table 6-6, cases (a) to (d). For the case of a simple parallel-tuned circuit, it can be noise-matched to give $A_n = 2$, which is the lowest figure possible. At the 3dB bandwidth the noise is exactly 1.5 times higher, with $A_n = 3$, which is a negligible increase. Attempting to increase the bandwidth by damping the antenna results in a significantly worse noise factor. For example, if we wanted to reduce the Q -factor from 100 to 10, and we noise-matched for this damped condition, then A_n would be 10 times worse. This, itself, may not be a problem if the amplifier noise power, P_n , is very low; but, significantly, the additional θ term in (6-28) ensures that the noise factor is ten times worse even if P_n is negligible.

6.3.4.2 Series-Tuned Antenna

Table 6-6, cases (e) & (f). The series-tuned antenna presents a significantly different case. At resonance, the noise is unaffected by the amount of damping, if the damping is large (i.e. the amplifier has a high input impedance). This is to be expected because the damping resistor is in parallel with the low reactance of the tuned antenna, which therefore dominates. If the antenna is noise-matched at the resonant frequency then, at the 3dB points, the A_n term rises from 2 to approximately θ^2 . This is a larger figure than the case for the tuned antenna, but the noise can be evened across the spectrum by noise-matching at the band edges to give $A_n = \theta\sqrt{2}$ at the band edges and $A_n \approx \theta/\sqrt{2}$ at the band centre. Importantly, if the amplifier noise power, P_n , is very low then the additional noise due to the increase in A_n may not be significant; and – unlike the case of the parallel-tuned antenna – there is no additional θ term in the expression.

Table 6-6, cases (i) & (j). A different picture emerges if the degree of damping is low (i.e. the amplifier has a low input impedance). The noise at resonance is higher, but it does not vary

as much over the band. This is to be expected if the dominant noise source is R_d . Of course, because R_d is low, the bandwidth is much lower.

Configuration	Conditions	Amplifier Noise Term	Source Equation
Parallel-tuned: Basic configuration Figure 6-7a	a) At resonance, i.e. $Q_0 \bar{\Omega} = 0$, with $Q_0 \gg 1$	$\left \frac{R_n}{R\eta^2 Q_0^2} \right $	(6-25)
	b) At -3dB points, i.e. $Q_0 \bar{\Omega} = 1$, with $Q_0 \gg 1$	$\left \frac{R_n}{R\eta^2 Q_0^2} \sqrt{2} \right \cdot \sqrt{2}$	(6-26)
Parallel-tuned: with damping resistor Figure 6-7b	c) At resonance, i.e. $Q_0 \bar{\Omega} = 0$, with $Q_0 \gg \theta$	$\left \frac{R_n}{R\eta^2 Q_0^2} \theta \right \cdot \theta$	(6-27), (6-28) <i>plus additional θ term as shown in (6-28)</i>
	d) At -3dB point(s), i.e. $Q_0 \bar{\Omega} = \theta$, with $Q_0 \gg \theta$	$\left \frac{R_n}{R\eta^2 Q_0^2} \theta \sqrt{2} \right \cdot \theta \sqrt{2}$	(6-27), (6-28) <i>plus additional θ term as shown in (6-28)</i>
Series-tuned and damped Figure 6-7d	e) At resonance, i.e. $Q_0 \bar{\Omega} = 0$, with $\theta \gg 1$	$\left \frac{R_n}{R\eta^2} \right $	(6-29), (6-31)
	f) At -3dB point(s), i.e. $Q_0 \bar{\Omega} = \theta$, with $\theta \gg 1$	$\left \frac{R_n}{R\eta^2} \frac{\sqrt{2}}{\theta} \right \cdot \theta \cdot \sqrt{2}$	(6-29), (6-31)
Untuned and damped Figure 6-7c	g) At low frequency with $\theta \gg X/R \gg 1$	$\left \frac{R_n}{X\eta^2} \right \cdot \frac{X}{R}$	(6-29), (6-32)
	h) At -3dB point, i.e. $X/R = \theta$, with $\theta \gg 1$	$\left \frac{R_n}{R\eta^2} \frac{\sqrt{2}}{\theta} \right \cdot \theta \cdot \sqrt{2}$	(6-29), (6-31)

Table 6-6 : Amplifier noise-matching term for different configurations of antenna.
When noise-matched, the $\|x\|$ term equals 2.

6.3.4.3 Untuned Antenna

Table 6-6, cases (g) & (h). The untuned antenna presents a similar case to the series-tuned antenna but, whereas the series-tuned antenna could be expected to be used at its 'centre frequency', this is not the case for the untuned antenna because the corresponding frequency is zero. It is convenient, for the untuned antenna, to describe the performance in terms of its reactance – we noise-match to its *reactance* at the frequency in use but, even so, A_n has a value of $2Q$, which means that if we desire a wideband response we must use an amplifier with a suitably low noise power.

6.3.4.4 Example

An example will serve to summarise the above discussion. Consider an antenna with parameters $L = 3.368\text{mH}$ and $R = 10\Omega$. We will evaluate its performance with each form of tuned circuit.

a) Parallel Tuned. Tuned with 4.7nF the antenna will resonate at 40kHz . It will have a Q -factor of about 85 and therefore a bandwidth of 470Hz . At the centre of the band $A_n = 2$, rising to 3 at the 3dB points. The noise-matched condition is $R_n = 7166 R\eta^2$.

b) Series Tuned. To give the same bandwidth as above, we would have to operate into a very low impedance, with a corresponding increase in the noise-matching term, A_n . There is no advantage to using this configuration.

c) Parallel Tuned and Damped. Suppose we wish to reduce the Q -factor of the antenna to 6, to give us a bandwidth of 6.6kHz ; an increase of $\theta = 14$ times. Regardless of the effect on A_n the additional θ term in (6-28) shows that this antenna will not have a desirable noise performance.

d) Series Tuned and Damped. To widen the bandwidth by $\theta = 14$ times we choose an appropriate value for R_d depending on the turns ratio of the matching transformer. We noise-match to the edge of the band, resulting in $A_n = 40$ at the band edges and $A_n = 10$ at the band centre – an increase on the un-damped figure, of course, but far preferable to the case of the parallel-tuned damped antenna. The noise-matched condition is $R_n = 10 R\eta^2$.

e) Untuned Antenna. To use the antenna untuned we must ensure that its 3dB point is at 43.3kHz , at which frequency $X / R = 92$, (because, from (a), $X / R = 1$ at 470Hz). We noise-match to the edge of the band, resulting in $A_n = 260$ at that point, and the noise will vary only moderately over the band of interest. The noise-matched condition is $R_n = 65 R\eta^2$.

6.3.4.5 Summary

In the above example, the parallel-tuned antenna was able to achieve a very low noise factor. Widening the bandwidth by using a series-tuned and damped antenna increased the noise factor, and the increase was lower than that for an untuned antenna. The value of A_n for the tuned/damped antenna was such that it could be countered by the use of a low-noise amplifier (see §6.4.2 below). We also saw that the matching condition depended on the type of tuning.

6.3.5 Comparison with Atmospheric Noise

The discussion on noise-matching is somewhat academic if there is a large contribution from the atmospheric noise. If we assume that the antenna has been designed so that it is 'large' – i.e. that the thermal noise is insignificant – then the amplifier noise contribution only has to be small in relation to atmospheric noise and can be considerably larger than the thermal noise without detriment to the operation. This is obvious from (6-22), and we can write

$$\frac{1}{2} \sqrt{\frac{\hat{P}_n}{4kT_0}} A_n < \max(1, F_k \Phi) \quad (6-33)$$

as a condition for the amplifier's noise contribution to be insignificant. Clearly a large value for A_n does not matter if \hat{P}_n is small or $F_k \Phi$ is large. Some BiFet op-amps have a very low noise power and we can infer that there are circumstances where noise-matching can be disadvantageous if the matching transformer introduces additional losses or distortion.

6.3.6 Increasing Bandwidth by Equalisation

6.3.6.1 Post-amplifier filtering

So far, we have concentrated on increasing the bandwidth of the antenna by damping the tuned circuit or operating it into a sufficiently high impedance. However we can achieve the same goal by equalising the signal – that is, by processing it using a filter that boosts the off-resonance frequencies. In the limit, a matched filter would eliminate the filtering imposed by the antenna altogether. Clearly a filter following the preamplifier will not alter the SNR at any frequency – it merely alters the signal bandwidth – and the salient point is how the SNR compares with one of the damping schemes we have discussed previously.

If we consider the parallel-tuned antenna, we can see from (6-26) that, if the antenna is noise-matched at the resonant frequency then, at all frequencies,

$$A_n = 2 + (Q_0 \bar{\Omega})^2 \quad (6-34)$$

At the frequency of the unadjusted 3dB bandwidth we have $(Q_0 \bar{\Omega})^2 = 1$. Now, if the filter boosts the off-resonance frequencies such that the 3dB bandwidth is increased by θ times; then $\bar{\Omega}$ will become $\theta \bar{\Omega}$ and so we will have

$$A_n = 2 + \theta^2 \quad (6-35)$$

This value for the noise-matching term is identical to that for the series-tuned and damped antenna. We can conclude that there is no particular advantage in using a filter to increase the bandwidth of a parallel-tuned antenna.

6.3.6.2 Magnetic Field Feedback

It is possible to design a filter for the forward signal path, with a suitably ‘inverse’ response to the antenna, using analogue circuitry, but there is the possibility that its response will drift. A simpler approach is to apply feedback to the system, by having the receiver amplifier generate a magnetic field in an inductor coupled to the antenna¹⁴. However, this feedback clearly cannot alter the SNR and it serves only to increase the signal bandwidth in exactly the same manner as the forward-path filter described above.

We can observe, in passing, that if the feedback were positive instead of negative then the effect would be to *reduce* the bandwidth. Radio Amateurs have made use of this ‘Q-multiplier’ scheme to increase the directivity of direction-finding antennas, where the broadband feedback circuitry is far easier to construct than a narrowband filter in the forward path.

6.3.7 Increasing Bandwidth by Antenna Design

We saw, in §4.2.6, that the bandwidth of a tuned antenna was proportional to its radius divided by its mass, or to the cube of its radius divided by the specific aperture squared. Given that mass and specific aperture are fundamental design parameters we can see that a simple way to increase the bandwidth of the antenna, whilst maintaining a specified performance, is to increase its physical size.

6.4 Amplifier Noise

The results of the previous section allow us to perform noise-matching, involving the tailoring of the turns ratio of the matching transformer to the noise characteristics of the amplifier. We must now consider these characteristics, and the design of a suitable preceiver preamplifier. Although our receiver may have to operate at up to 200kHz, we can still employ design techniques that are used for audio equipment. Some design tips for audio pre-amplifiers were given in (Adam, 1989); and some guiding principles for low-noise audio design were given by (Paschal, 1988). Paschal discusses a discrete transistor design which can be updated to use modern ICs, but his discussion of matching transformers is more relevant, given that the prevalence of high-performance ICs now makes the design process much easier.

6.4.1 Discrete Devices

One of the most demanding low-noise, low-distortion tasks in audio engineering is the design of a pre-amp for a moving-coil cartridge in a record player. Until recent years, op-amps were too noisy to be considered for this task, and equipment usually made use of a transformer

¹⁴ I am grateful to Prof. E. Vilar for drawing this to my attention.

or a pre-amp constructed from parallel-wired transistor stages. A number of discrete designs evolved and we will review some briefly, for historical interest.

One of the most interesting discrete designs was due to (Nordholt and Van Vierzen, 1980), and reproduced in Elektor magazine (Anon., 1980) where it was shown that the noise performance of a simple transistor amplifier is dependent on the internal base resistance, r_b . A device was identified with a particularly low value for this parameter, and a three-transistor amplifier constructed with a measured noise voltage floor of under $0.5\text{nV}/\sqrt{\text{Hz}}$ (approximately the noise of a 10Ω resistor) and a total harmonic distortion at 100Hz of 0.017% (-75dBc).

An equally elegant design, and with a more impressive performance, was described by (Self, 1987) in which the low-noise input transistor was followed by an op-amp, for gain. To avoid the +0.6V offset from the first stage leading to a wholly impossible offset at the output, the op-amp was provided with a feedback loop to stabilise its d.c. operating point; this being preferable to a.c. coupling where a response down to 20Hz is required.

The traditional solution to the d.c. offset problem is to use a differential pair as the input. Many such designs exist, where a low-noise transistor is followed by an op-amp. Typical amongst them is one from Analog Devices (Anon., 1992), which uses a monolithic transistor pair driving a – for its time – industry-standard low-noise op-amp. A noise of $0.5\text{nV}/\sqrt{\text{Hz}}$ is claimed, and a gain of 1000 over a 200kHz bandwidth.

6.4.2 Integrated Circuits

The state of the art of monolithic integrated circuits is such that convenient op-amp building blocks are now available ‘off the shelf’ and we will probably not need to resort to a discrete transistor design. Some IC op-amps – none of them particularly ‘modern’ devices – are listed in *Table 6-7* below. We will not discuss these devices in any detail except to say that there is clearly a large difference in the way these devices would be used.

Device	Noise voltage, [nV/ $\sqrt{\text{Hz}}$]	Noise Current [pA/ $\sqrt{\text{Hz}}$]	Noise Power [zW/Hz]	Noise Resistance [k Ω]	$P_n / 4kT$
AD797	0.9	2.0	1.8	0.45	0.11
AD8055	6	1	6	6	0.38
AD9632	4.3	2	8.6	2.2	0.54
CLC425	1.05	1.6	1.7	0.66	0.11
TL071	18	0.01	0.18	1800	0.011
TLC2201	8	0.0006	0.0048	13300	0.0003

Table 6-7 : Noise characteristics of some operational amplifiers.
This table is only a summary; the variation of noise with frequency is important, as is the distortion.

For example, the BiFet TLC2201 device has a very low noise current and so it is suited for high-reactance untuned antennas, but its frequency response is such that its use is restricted to a few kHz, for example in a radiolocation beacon. With this device, noise matching will probably not be required because its noise power is so low.

The AD797 was the first of a range of very high performance devices that are characterised by the fact that their noise performance is maintained up to and beyond 100kHz. Most 'traditional' op-amps have a very poor performance outside the audio band, due to the internal design of the gain stages. The AD797 also maintains a very low distortion and certainly refutes the traditional view that discrete designs are needed for good noise and distortion performance.

In recent years companies such as Analog Devices have specialised in an ever increasing range of high-performance op-amps, featuring video bandwidth, ultra-low distortion and ultra-low noise, with – importantly – these characteristic maintained at high frequencies. Where some of these device fail to perform is that the open-loop gain presents a low 3dB point. For example, the AD9631 has a noise power of¹⁵ 8.6zW/Hz, a noise resistance of 2.2kΩ and a distortion of -120dB (all characteristics flat to 300kHz) but it has an open-loop gain of only 1000 at 100kHz. This is only a rough figure, of course, but it means that there is little headroom for us to define a precise closed loop gain, which we might wish to do.

We have examined several op-amps and decided to use the AD8055 in the wideband channel sounder receiver. This is mainly because it is cheap and, being a voltage feedback device, easier to work with than some of the modern current-feedback devices. It has a noise power of 6zW/Hz and noise resistance of 6kΩ. The distortion is -100dB at 100kHz and the gain is 4000 at up to 1MHz.

6.5 Interference

Interference can be taken to include co-channel and adjacent channel broadcasts, as well as impulsive noise due to atmospheric and man-made disturbances, and – at the lower end of the spectrum – interference from the national grid at 50Hz and its harmonics. The LF band contains a good deal of telemetry traffic and some powerful transmitters, including, for example, a military installation at around 21kHz, which interferes with commercial location equipment; and the Loran navigation beacons in the region of 90–110kHz which can cause serious interference to low-frequency sub-surface communications. A number of methods of interference cancellation have been proposed by the author and others. Some of these methods are also appropriate for reducing broadband atmospheric noise, and we will briefly review them here.

¹⁵ The SI prefix *zepto*, symbol 'z' indicates a unit multiplier of 10^{-21} .

6.5.1 Shielding

Atmospheric-borne interference has its worst effect on the surface, where the signals from an underground antenna are weak, so one method, proposed by (Drummond and Putnam, 1998) is simply to site the ‘surface’ receiver underground where it is shielded from the full effects of atmospheric noise.

Since the atmospheric magnetic noise signal is attenuated at only 8.7dB per skin depth, the improvement is likely to be minimal for a receiver that is only a short distance below the surface. However, Drummond and Putnam did report some success, so it may be that, above ground, their unshielded receiver was detecting the electric noise field by electrostatic coupling as discussed in §6.2.6.

An electric field is attenuated by a factor of $\sigma / \epsilon_0\omega$ (provided that $\sigma / \epsilon\omega \gg 1$) as it passes the air/rock boundary; and with typical values of conductivity $\sigma = 1\text{mS/m}$, permittivity $\epsilon = 7\epsilon_0$ and frequency 100kHz, the cross-boundary attenuation of the electric field would be 28dB. This would account for a significant improvement in SNR, but one that could also be achieved by adequate electrostatic shielding and balancing of the receiver loop.

Another result of the rock being a good conductor, with $\sigma / \epsilon\omega \gg 1$ is that the rock has a high refractive index, and the electromagnetic field therefore propagates in a direction almost normal to the interface. Since the **E** and **H** vectors are normal to the direction of propagation this may mean that a horizontal loop antenna picks up only a small amount of electromagnetic noise. Compared with a vertical loop antenna. The precise situation clearly depends on the local geology, and a systematic investigation of underground noise would appear to be desirable.

6.5.2 Using a Large Transmitter

It was noted, earlier, that a receiver antenna has a ‘maximum useful size’ – if it exceeds this size it merely picks up more noise, and the SNR remains the same. However, the SNR can be improved by using a larger transmitter. If, for the same mass, we double the radius, then the specific aperture will double and the SNR at the receiver will increase by 6dB. Doubling the mass or the power consumption will result in only a 3dB increase in SNR. The need for a large transmitter may be particularly important for the up-link where, unfortunately, it may not always be possible to utilise a physically large antenna. Earth current systems can help here, because they provide an easy way to make a large underground antenna.

6.5.3 Positioning to minimise noise

If the noise source is directional, then positioning the receiver to minimise the noise can be more successful than positioning it to maximise the signal (although there is some practical

evidence to support our earlier conclusion that a coplanar orientation resulted in an increased signal strength at larger distances). Of course, antenna orientation may not help if the interference is elliptically polarised, and setting up an induction loop system with tilted loops requires operators with a certain amount of skill. A method of antenna-orientation for an earth-current receiver, implemented by switching different sets of electrodes, was described by (Fauchez, 1998)

'Null-steering', a tool often used at r.f., would appear to be a concept that could be used with induction radio. Two or three orthogonal loop antennas (or two orthogonal earth-current dipoles) could feed a DSP-implemented steerable phase-shifter.

6.5.4 Noise-Cancelling Antennas

6.5.4.1 Field Gradient Method

Perhaps the obvious method of reducing noise and interference is to use a field gradient technique similar to that adopted in instruments such as magnetic and gravity gradiometers. Two co-axial induction loops, with their axis aimed at the transmitter, will detect slightly different magnetic fields. By connecting the loops in anti-series we can detect the difference between the two signals. External sources of interference will induce a similar signal in each loop, which will cancel. If we consider the axial magnetic field from an induction loop antenna to obey an inverse cube law, of

$$H = \frac{M}{2\pi d^3} \quad (6-36)$$

and we utilise a pair of receiver loops at a distance of $d \pm \frac{1}{2}\Delta d$, then the effective field we detect can be shown to be

$$H_{\text{eff}} = \frac{M}{2\pi d^3} \frac{3\Delta d}{d}, \quad \text{with } \frac{\Delta d}{d} \ll 1 \quad (6-37)$$

We can see that the inverse cube law has become an inverse fourth power, and that for a typical situation of $d = 100\text{m}$ and $\Delta d = 1\text{m}$ the signal is reduced by 30dB. Clearly, for this scheme to be feasible we must ensure that the interfering signal is cancelled to at least this degree, which would require careful attention to the antenna design and we would also have to analyse the effect of mutual coupling between the receivers.

6.5.4.2 E-field Cancellation

A similar scheme to the above, suggested by C. Trayner (*pers. comm.*), takes advantage of the fact that far-field interference maybe mainly electromagnetic in nature, whereas an induction radio signal is mainly magnetic. A whip antenna will mainly pick up just the *E*-field interference, which we assume will be in a fixed proportion to the *H*-field interference. The

presumption is that the whip will detect only a very small signal from the induction radio, thus providing a method of isolating the interference from the wanted signal. Unfortunately, this method relies on the electric field of the interference having a fixed relationship to the magnetic field and, as we explained in 6.2.1, this will not be the case.

However, where the interference is predominantly co-channel or adjacent-channel broadcasts, rather than man-made or atmospheric noise, then the technique may have some merit. The method would not suffer from the loss in signal strength of the field gradient method, but it is likely to require DSP techniques to control the phase and amplitude of the interference to obtain adequate cancellation in the receiver. A self-learning routine, with the induction loop transmitter broadcasting a known pilot tone or sequence could perhaps be utilised. This method should work with induction loops, but it might not be so successful with earth-current systems because these generate and detect an electric signal component, as well as the magnetic component.

6.5.5 Spatial Filtering

An AM communications receiver design by Trevor Brook of Surrey Electronics (Brook, 1989) provided a number of demodulation methods. As well as the conventional ‘envelope’ detector output, and a synchronous demodulator for double and single sideband signals, the instrument allowed the user to demodulate with a quadrature carrier, which has the effect of nulling the strongest station, and it also featured an ‘independent sideband’ (ISB) or ‘stereo spread’ mode where the lower sideband was placed on one channel and the upper sideband on the other channel of a stereo pair. The receiver architecture is shown in *Figure 6-8b* on page 162, where it can be seen to be an extension of the architecture we would normally use for SSB reception (*Figure 6-8a*).

Listened to on stereo headphones, this technique is known as ‘spatial filtering’, or ‘binaural filtering’ where different signals appear to come from different positions in the stereo image. Brooks noted that there were “interesting effects on fading signals and where the two sidebands are suffering different interference”. He also made the unexpected observation that “propagation effects cause distant lightning static to sound quite different in the two sidebands”.

(Fauchez, 1999) explained an interesting additional mode of operation where in-phase and quadrature demodulation of the *same* sideband spreads it over a stereo image to give what Fauchez termed a “voice of God” characteristic caused, we may surmise, by the complementary phase response of the outphaser circuits.

6.5.6 Adjacent Channel Cancellation

Adjacent channel cancellation uses similar equipment to that used for spacial filtering, and can be considered to be an extension to that method. Methods proposed by (Taylor, 1977) and (Illingworth, 1981) both rely on properties of AM or double-sideband signals and the methods were proposed as a means of improving broadcast reception in the medium wave and short-wave bands.

6.5.6.1 Carrier extraction and demodulation

Before implementing the method of interference cancellation, it is necessary to extract or synthesise the carrier of a DSB or AM signal. This can be done in a Costas Loop, as shown in *Figure 6-8b* or in a squaring loop or, in a somewhat complicated fashion, using a frequency-locked loop (FLL).

Phase-locked loops (such as Costas') do not perform well under conditions of high interference. Some dynamic adjustment of PLL bandwidth (Brook, 1989) helps, but an alternative solution is to use an FLL (Illingworth, 1981). A VCO is used to extract I and Q channels, as with the Costas loop, but these low-pass filtered outputs are fed to a frequency discriminator to drive the VCO. The result is that the loop tracks frequency but not phase. Since a phase-coherent output is essential, the error signals are re-modulated by the VCO and summed to reconstruct a new 'clean' carrier, which can then be used for coherent I and Q demodulation.

Having derived the in-phase (I) and quadrature (Q) carriers, these are used to synchronously demodulate the AM/DSB signal, generating an interference component that can be further processed. The technique described so far is that of a conventional homodyne receiver, as shown in *Figure 6-8b*. The I channel contains the signal plus the interference, whereas the Q channel contains the quadrature interference only.

6.5.6.2 Interference cancellation

If the interfering signal is in only one sideband then the interference in the 'Q' channel is phase-shifted by 90° and used to cancel the interference in the 'I' channel. This is tantamount to having a sideband selection switch, as shown in *Figure 6-8b*. Note that the sense of the cancellation depends on whether the interference is in the LSB or USB, so this method cannot cope with interference that spans both sidebands.

For interference that is an AM signal, Taylor's technique is to lock onto the interfering carrier that is present in the Q channel, which is then re-modulated using a digital waveform with a particular mathematical property, such that it can be used to cancel the interference in the 'I' channel. This method works even if the interfering AM signal overlaps both sidebands.

For interference that does not fall into either of the above categories, Illingworth's technique used a further modulation to derive a low-frequency term which can be used, dynamically, to determine the sense of the error correction required – this can be altered, as required, at a high audio rate. Illingworth described two slightly different methods – the choice depends on the nature of the interference. All these methods can be described diagrammatically by adding a 'black box' to the receiver architecture, as shown in *Figure 6-8c*.

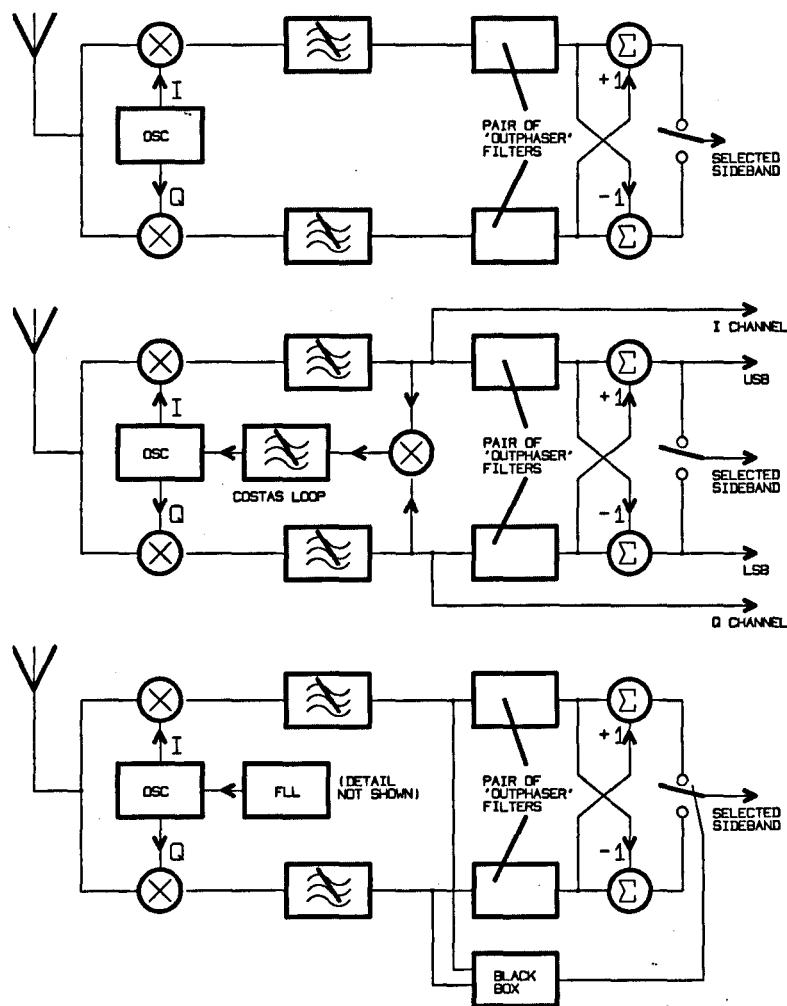


Figure 6-8 : Receiver architectures for interference rejection.

Top: a) a standard 'outphaser' SSB radio allows the selection of USB or LSB. This design will receive a DSB signal, but only one sideband can be listened to, because oscillator drift will cause an unpleasant beat.

Middle: b) Adding a Costas Loop allows the local oscillator to be synchronised to a double-sideband input. USB and LSB outputs can now be listened to as a stereo pair – or a single sideband can be selected as before. DSB, I & Q channels are also available.

Bottom: c) Adding a suitable 'interference processor' black box (see text) allows us to dynamically select the 'best' sideband (at a rate that can be well above audio) in order to completely remove an interfering signal that spans both sidebands. A frequency-locked loop (FLL) allows greater immunity to interference but it is more complicated to implement.

6.5.7 Quadrature Phase Space

LF broadcast stations are designed to be received using simple envelope demodulators which are not phase sensitive. We can therefore deduce that the spectrum must be relatively quiet at this point *apart* from the AM broadcast signal. We have proposed that the quadrature phase space of a broadcast station would therefore be a suitably quiet place to operate an induction radio system (Gibson, 2002).

Consider a single-sideband transmission that is frequency-locked to an AM broadcast carrier. The local SSB signal, from a small induction loop, will be undetectable by other users of the spectrum unless they are in the immediate vicinity of our transmitter¹⁶. At the receiver, we synchronously demodulate the SSB transmission in quadrature to the AM carrier. The AM sidebands will be ‘invisible’ to this demodulation, although we may detect the 25Hz phase-modulated Radio Data System signals¹⁷, which are easily filtered out. The background noise in the absent sideband will overlie our demodulated signal, reducing the SNR by 3dB. Unfortunately, we cannot perform a complete SSB demodulation (e.g. see (Gibson, 1996a; b), which would eliminate this noise, because to do so would require an in-phase demodulation, and our signal would be swamped with the AM sidebands.

Alternatively, consider a double-sideband suppressed carrier (DSB-SC) transmission, synchronised in quadrature to the AM broadcast carrier. The combination of a DSB-SC signal; and a quadrature carrier synthesises a narrowband FM signal and so this signal will not be detectable to a local user with a conventional AM receiver. The issue of secure communications frequently arises during cave rescues, which is one reason why telephone systems may be used in preference to radio communications. At the receiver, we synchronously demodulate the DSB-SC transmission in quadrature to the AM carrier. As before, the AM broadcast sidebands will be ‘invisible’ to this demodulation.

There is a potential difficulty with a DSB-SC system when we consider the up-link communications. The local carrier, used as the reference for the underground transmitter will be delayed in phase by its passage through the rock. In a similar fashion, the up-link signal will suffer a phase delay. Thus the magnitude of the demodulated signal will be proportional to $\cos\phi$ where ϕ is the total phase delay through the rock. If we assume that the maximum acceptable delay is 60° , corresponding to 6dB of attenuation, then we are limited to a depth of $\lambda/12$. In rock, the wavelength λ is 2π times the skin depth δ , so we are limited to around half a skin

¹⁶ Notably, the UK licence regulations (MPT 1337) permit induction loop operation at up to 185kHz., overlapping with a number of long wave broadcast stations.

¹⁷ RDS on BBC Radio 4, long-wave, consists of binary data at 25 bit/s, biphase-modulated onto a 25Hz binary carrier that then phase-modulates the 198kHz carrier at $\pm 22.5^\circ$. The combination of low data rate and low modulation angle ensures that there is no interference with the signal. See (Anon., 1993).

depth. This problem could be eliminated by altering the reference phase of the underground transmitter. This would require some dialogue between the transmitter and receiver but, as part of an adaptive system architecture, this would be trivial. In any case, the phase only needs to be accurate to perhaps 30°.

DSB-SC transmissions are usually avoided in simple analogue equipment because of the additional complexities of demodulation, which requires that an accurate carrier be reconstructed at the receiver. If the local carrier is close in frequency, but is not phase-locked then we suffer a 3dB penalty in SNR. AM and NBFM are avoided because of power considerations. In the above scheme we have, essentially, 'hijacked' a carrier so that we can transmit DSB-SC and yet demodulate it as if we had transmitted an AM or NBFM signal. Use of the supposedly quiet quadrature phase space will allow us to investigate areas of the spectrum that were previously thought unavailable for sub-surface use. Therefore this is an area of research that would be worthwhile pursuing.

6.5.8 Speech Processing

The above methods have concentrated on the elimination of interference, but the signal to noise ratio can also benefit from suitable signal enhancement. In digital systems this may take the form of coding and error correction algorithms; and in analogue speech-based systems we may make use of pre-emphasis or other modification to the signal bandwidth.

It is also common to reduce the dynamic range of an audio signal, in order to better utilise the available transmitter power, but many methods of achieving this achieve only a modest increase in SNR. However, a study by (Thomas, 1976), which was reported by (Dance, 1985), claimed that an advanced form of level control could 'punch up' the average loudness of a communications system by as much as 20dB without exceeding peak modulation levels. It was also claimed to increase the mean power level by 6dB, and by 15dB on speech consonants, which are critical for intelligibility. In 1976, digital speech processing was slow and expensive, but we can now re-assess Thomas's system using simple low-cost circuitry. We described this in (Gibson, 1999) and feel that further work in this area could be highly beneficial.

6.6 Concluding Remarks

As with the case of transmitter antenna, the main design choices for the receiver antenna are tuned v. untuned and ferrite v. air-cored. In this chapter we have concentrated on the distinction between tuned and untuned antennas, although this cannot be considered in isolation to the other criterion because a ferrite-cored solenoid is likely to have a very low bandwidth when tuned, and would therefore have to be untuned for speech applications.

We showed that the noise performance of the system could be described by the noise factor, with terms representing atmospheric and amplifier noise. We noted that the CCIR/ITU atmospheric noise temperature data, which was collected using an electric antenna, could not be used to derive the noise received by a magnetic antenna; a point which highlights the importance of the noise data we wish to collect using our channel sounder.

Regarding the amplifier noise, we discussed how this could be reduced by attention to noise-matching of the amplifier to the antenna; and how this depended on the type of antenna in use. We saw that the performance of modern operational amplifiers is such that we do not need to consider discrete designs for low-noise pre-amplifiers. As with other aspects of this analysis, it is not possible to draw definite conclusions on the appropriate device to use without a system specification – in particular the amplifier of choice will depend on the frequency we wish to use and the type of antenna.

We looked at a number of methods of mitigating external noise and interference, all of which would bear some further investigation. In particular we described a novel technique of utilising the ‘quadrature phase space’ underneath a broadcast carrier.

There will always be a limit on the useful size of the receiver antenna, which is governed by the background noise levels but, often, the system will benefit from the transmitter antenna being as large as possible. Through-rock communications systems that required only a ‘down-link’ transmission could utilise a physically large underground receiver. However, for an up-link transmission, the surface antenna might not benefit from being as large. Therefore, we can re-iterate a remark we made at the end of the last chapter, that not only could the transmitter and receiver antenna benefit from different designs, but the up-link and down-link may also require different designs.

The information presented in this chapter can be used in the design of a receiver antenna that is tailored to a particular communications system. In particular we can make use of the untuned antenna analysis in the design of a wideband antenna suitable for channel sounding. We are now in a position to discuss channel sounding and to apply the results of the last two chapters to the design of a channel-sounding system.

6.7 References¹⁸

- Adam, W. (1989). Designing Low-Noise Audio Amplifiers. *Electronics & Wireless World* 95(1640), 628-633.

¹⁸ References to *CREGJ* are to the *Cave Radio & Electronics Group Journal*; a non-peer-reviewed low-circulation magazine published by a special interest group of the *British Cave Research Association*, bcra.org.uk. The CREG journal is filed at the British Library as ISSN 1361-4800.

- Anon. (1980). Moving-Coil Preamp. *Elektor* 6(6), July/August 1980, 7-98.
- Anon. (1992). Application Note AN136: An Ultralow Noise Preamplifier. In *Audio/Video Reference Manual*. ed.: Analog Devices, pp. 11.81-11.82.
- Anon. (1993). Low-Cost 198 kHz Radio Data Receiver, Application Note AN86, Gec-Plessey Semiconductors. *Electronics World* 99(1692), 960-1.
- Brook, T. (1989). A.M. Synchronous Demodulator. *Electronics & Wireless World* 95(1643), 858-60.
- Burley, S., M. Darnell, C. Prowse and R. Chandler (1997). *Improved Detection in Correlated Impulsive Noise Using Wavelet De-Noising*. Proc. IEEE Int. conf. MILCOM 97, Monterey USA.
- CCIR (1968). *Report 322. World Distribution and Characteristics of Atmospheric Radio Noise*. Xth Plenary Assembly. Comité Consultatif International des Radiocommunications, Geneva.
- Dance, B. (1985). Digital Speech Processing. *Practical Wireless*, April 1985, 20-22.
- Drummond, I. and I. Putnam (1998). Going Deeper - Two Ways to Improve Performance. *CREGJ* 34, 24-25.
- Fauchez, J.-J. (1998). The Berger on 137 kHz. *CREGJ* 34, 12.
- (1999). Spacial Filtering as an Aid to Signal Reception. *CREGJ* 37, 15-16.
- Gibson, D. (1996a). Designing an SSB Out-Phaser (Part 1). *Electronics World* 102(1721), 306-310.
- (1996b). Designing an SSB Out-Phaser (Part 2). *Electronics World* 102(1722), 392-394.
- (1999). *Low-Cost Implementation of a Speech Compression System*. Proceedings of 5th International Symposium on Digital Signal Processing for Communications Systems (DSPCS '99), Perth-Australia.
- (2002). Cave Radio Notebook 51: Cave Radio Using a Broadcast Carrier. *CREGJ* 49, 23-24.
- Illingworth, L. (1981). A.M. Receivers without Interference. *Wireless World* 87(1549), 79-83.
- ITU (2001). Recommendation ITU-R P.372-7. Radio Noise. Geneva, International Telecommunications Union.
- Nordholt, E. H. and R. M. Van Vierzen (1980). Ultra-Low-Noise Preamplifier for Moving-Coil Phono Cartridges. *Journal of the Audio Engineering Society* 28(4), 219-224.
- Paschal, E. W. (1988). *The Design of Broad-Band VLF Receivers with Air-Core Loop Antennas*. Anacortes, WA: Whistler Radio Services. (2nd Edition).
- Rabson, J. (1993). Expected Values of Atmospheric Radio Noise. *CREGJ* 14, 26.
- Ramo, S., J. R. Whinnery and T. Van Duzer (1984). *Fields and Waves in Communication Electronics*. New York: John Wiley. (2nd Edition).
- Self, D. (1987). Design of Moving-Coil Head Amplifiers. *Electronics and Wireless World* 93(1622), 1206-1209.
- Taylor, P. L. (1977). Eliminating Adjacent-Channel Interference. *Wireless World* 83(1499), 55-57.
- Thomas, L. D. (1976). An Investigation into a Non-Linear Method of Automatic Level Control for Speech Channel Signals (Msc Thesis). Swansea: University College of Swansea.
- Watt, A. (1967). *VLF Radio Engineering. [International Series in Electromagnetic Waves: 14]*. Oxford: Pergamon Press.

7 Channel Sounding

Abstract: A wideband channel sounder is described that uses a near-pseudo-random sounding sequence and extensive signal averaging at the receiver. In the context of through-rock sounding, ‘wideband’ means a frequency range of at least 4:1, so the design of the magnetic dipole antennas – which would normally have a high Q-factor, is important. The wideband sounder is essentially spread-spectrum device, and the low signal/noise ratio at the receiver means that a code-locked loop must be used to synchronise the systems. However, the sounding sequence has, potentially, a non-ideal auto-correlation function because it is not quite a maximal-length sequence, and so the sequence design has also to be addressed.

We noted, in the discussion on propagation in chapter 2, that there existed an optimum frequency for communications that depended on the required communications range and the characteristics of the medium; however, we might not be able to use this optimum frequency in the presence of background noise and co-channel interference. To lessen the effect of such interference we can either choose a different frequency, or we can evaluate a different orientation of transmitter and receiver; and we have noted, briefly, that the optimum frequency varies with antenna orientation. Taking these factors in to account we can see that there may be advantages in utilising an adaptive communications system. This will require some degree of dynamic channel evaluation, which we shall discuss in this chapter.

7.1 Principles of Channel Sounding

The procedure of evaluating a communications path is called *channel sounding*. For a simple implementation, we could connect a variable-frequency signal generator to an induction loop transmitter, and make individual signal-strength measurements using a tuned receiver but, especially if we wish to evaluate the channel dynamically, a swifter, automated exercise would be preferred.

7.1.1 Purposes

There are three purposes for which we need a method of channel-sounding,

- i) to evaluate the propagation of the field in different conditions (e.g for investigation of optimum frequency)

- ii) to record the background noise and interference
- iii) to dynamically evaluate the channel, as part of an adaptive communications system.

The first two purposes can contribute towards a general evaluation of the spectrum, from which we would hope to draw some conclusions on the optimum frequency to use. For example, cave rescue communications equipment in the UK currently uses USB transmission on 87.5kHz. This can suffer from interference from the *Decca LORAN* beacons.

Noise evaluation at VLF using a sampled-data system has been described by (Laflin, 1992) who used a calibrated antenna and a digital tape recorder, sampling at 48kHz, to record urban noise. Laflin used a minimum 4-term Blackman-Harris time domain window with Fourier transform sample lengths of 32768 and 65536 samples. His antenna and pre-amplifier presented a 3dB bandwidth extending from 10Hz to 10kHz and a response of 50mV/nT. These figures can be compared with those given in §9.3.2 and §9.3.3.

Channel sounding – especially if it is a swift, automated exercise – will allow us to evaluate radically different frequencies, and different antenna configurations. We also envisage operating the channel sounder as a field data logger to capture and evaluate noise and propagation over a period of many months.

Some further purposes of channel sounding are

- iv) To find the impulse response of the medium – this may be useful for modelling purposes.
- v) To investigate wideband systems and frequency agile systems – are they feasible, given the noise and interference spectrum?
- vi) To investigate near and far fields, and secondary fields for modelling purposes – e.g. we could implement an *E*-field receiver to measure *E / H* field ratio, and we could use three orthogonal antennas to look at secondary field characteristics.

7.1.1.1 Frequency Stepping

An induction loop antenna usually has a high *Q*-factor and, unless it is tuned to resonance, it is very inefficient. This poses problems for wideband channel sounding, but we could envisage a frequency-stepping system with some method of re-tuning the antenna between measurements. We saw, in chapter 5, that the tuning capacitor has to withstand large stresses. A tuning system would therefore have to comprise either a bank of high-performance capacitors to be switched when required (either manually or via electromechanical relays) or an air-vaned variable capacitor (perhaps with a servo drive) that was rated for the necessary voltage and

current. Such devices are available from specialist suppliers, but they are expensive and bulky (typically 300mm × 100mm × 100mm).

Alternatively, we could design the antenna to have a low Q -factor, by constructing it with a large radius. This may be preferable, but the system would be less portable, and additionally, it could approach radiation resonance at the top end of the spectrum, and would not behave as a point source if the field point were too close.

The tuning problems apply also to the receiver. Readjusting a transmitter and receiver to a new frequency takes time, even if done automatically; and for a large number of data samples, a spot-frequency system could be slow to operate.

7.1.1.2 *Wideband Methods*

A wideband system is inevitably speedier to operate than a frequency-stepping system. The transmitter generates a short burst of wideband energy, which is captured by a wideband receiver incorporating a high-speed sampler. We then perform a digital signal analysis to extract the data. Additionally, if such a sounder were used without a transmitted signal, it would function as form of spectrum analyser, and would allow us to investigate the background noise.

The problem with a wideband design is that the transmitter antenna will have to operate untuned, which reduces its efficiency. In addition, the PA will be inefficient but, as we saw in chapter 5, we can avoid this by transmitting a digital binary signal. We still have the problem of a low power level, but the transmitter PA can operate at a far higher efficiency than would be possible if we were to transmit an analogue signal. Several wideband methods could be considered,

- i) Digital sequences, such as pseudorandom binary (PRB)
- ii) Chirp methods (FM, frequency sweeping)
- iii) FFT-derived waveforms

All have the advantage that because they are periodic, some degree of signal-averaging is possible at the receiver, to increase the SNR. However, only the first method uses a digital sequence, which is necessary if we are operate at a high PA efficiency.

7.1.1.3 *Signal Averaging*

The inherent low power density of a wideband system means that the signal to noise (SNR) ratio at the receiver is likely to be very low. We can increase this by transmitting multiple copies of the sounding sequence and averaging the results – if we transmit N sequences then the signal/noise voltage ratio will increase by \sqrt{N} , assuming the noise is uncorrelated and that we

can adequately synchronise the receiver to the transmitter. We could achieve the same effect by boosting the transmitter power but this may be limited by the equipment design. The advantage of signal averaging is that we can choose to average any number of datasets, as circumstances dictate, without the problem of a power limitation at the transmitter.

In a digital system it can be more appropriate to discuss the *bit error rate* (BER) than the SNR, and we know from standard data communications theory that the bit error rate is related to the ratio E/N_0 – where E is the energy per bit and N_0 is the noise power spectral density. Clearly, to achieve a high E we can utilise a high power for a short time, or a low power for a longer time – for signal averaging we choose the latter but, additionally, we split the energy into non-contiguous fragments in order to preserve the frequency spectrum of the transmitted signal.

7.2 Outline of Proposed Method

We will now outline a proposed method of wideband channel sounding. A block diagram showing the is in *Figure 7-1* below.

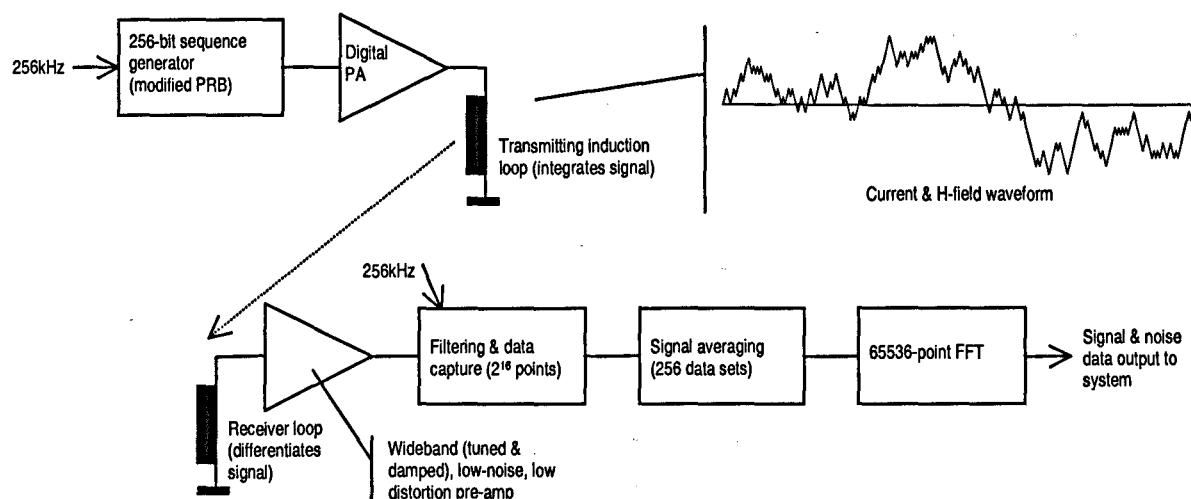


Figure 7-1 : Block diagram showing intended operation of the channel sounder.

7.2.1 Sequence Design

7.2.1.1 M+1 Sequence

We have seen that it will be advantageous to transmit a binary sequence, and the obvious sequence, to generate a wideband spectrum, is a maximal length pseudo-random (PRB) sequence. This is problematic, because binary m -sequences are of period $2^n - 1$ (for some integer n) and thus cannot be used directly with the most efficient form of fast Fourier transform (FFT), which requires a sequence of period 2^n . Two well-known solutions to this problem are

discussed by (Wellstead, 1975), namely i) to ignore the problem, and to proceed as if the continuous signal has a period of 2^n ; and ii) to use the inverse Fourier transform to generate the test data itself – the technique mentioned in §7.1.1.2, (iii) above. In the latter case, the resulting test signal will have the prescribed length and power spectrum, but will, unfortunately, consist of non-binary values whereas, for PA efficiency, as we have already stated, we need a binary sequence, so that we can use a digital or class-D power amplifier.

The solution we have adopted is to take a maximal-length PRB sequence and add one bit, to make a sequence of length 2^n . The position of the added bit affects the size of the sidelobes in the auto-correlation function (ACF) but, by trial and error, it is possible to find the position that results in the lowest sidelobes. Clearly, the precise power spectrum of the sequence does not matter for simple spectral measurements of attenuation and phase shift, but for impulse response measurements we will need to calculate the inverse to the sequence, as we will discuss in the next chapter.

For sounding purposes we do not need a particularly fine resolution; for example, a 256-chip sequence, clocked at 256kHz would result in spectral lines 1kHz apart, which is probably adequate for measuring how the propagation varies with frequency.

7.2.1.2 *Synchronisation*

We are envisaging a SNR that is low enough to require signal-averaging at the receiver. Suppose we transmit a sequence of 256 chips, and we repeat it 65536 times. For signal averaging to present its full advantage the receiver must be able to maintain synchronism to the transmitter to perhaps 10% of a chip width over the time taken to transmit all 65536 copies of the sequence, which requires a phase accuracy of around 0.006ppm. Clearly it will be necessary to provide some method of synchronising the receiver to the transmitter, as this level of accuracy could not easily be achieved otherwise.

During the early stages of this project, we considered the use of a cable connection between the two devices, perhaps a fibre-optic cable to avoid questions of signal conduction along the wire. However, this scheme would not be fully portable. We also considered synchronisation to the BBC R4 broadcast station on 198kHz or to the time signal transmission MSF Rugby on 60kHz, both of which are detectable underground. Although this method would not provide a solution to the problem of system identification and impulse response measurement, it might have provided a quicker development route than the scheme we decided to adopt.

The need for signal averaging suggests that the transmitted signal – being in the nature of a spread-spectrum – will be below the level of the background noise, so we are not able to use a conventional phase-locked loop, but we can use a variation on the standard code-locked or

delay-locked loop. Our problem is now that the modified $M+1$ sequence we wish to use does not have a very good auto-correlation property.

The need to synchronise the receiver to a non-ideal sequence has lead us to a detailed investigation of the design of sequences for communications system identification. This topic will be explained in depth in the next chapter. For the time being, it is only necessary to state that we will use a code-locked loop to extract a system identification signal with which we can synchronise the receiver to the transmitter; and an added bonus is that we can thereby obtain the impulse response of the channel.

7.2.2 Frequency Range

Conventionally, we may understand the term 'broad-band' to mean 'low Q ' but even a Q -factor of 10 only gives us a frequency range of about 1.1:1. For speech communications we may wish to evaluate the spectrum over a 4:1 range from about 45kHz to 180kHz. For beacon and radiolocation work, we might wish to assess frequencies from 500Hz to 8kHz – a range of 16:1. If the frequency range is we require is Ω^2 and the acceptable signal level at the band edges, relative to the centre, is A then the required Q factor (if we assume a simple first order tuned circuit) can be found from:

$$A = \frac{1}{\sqrt{1 + Q_0^2 \left(\Omega - \frac{1}{\Omega} \right)^2}} \quad (7-1)$$

where Q_0 is the Q -factor at the centre of the band.

This expression shows that for a frequency range of 4:1 and a band-edge attenuation of 3dB we need a Q -factor of exactly 2/3. With such a low value, we must be sure to use the correct design equations, many of which were approximated in chapter 6 with the assumption $Q \gg 1$. *Table 7-1*, below, presents some further relationships between F and Q .

Frequency range, Ω^2	Q factor for band-edge attenuation of		
	3dB	6dB	12dB
2.6:1	1.01	1.74	3.88
4:1	0.67	1.15	2.57
10:1	0.35	0.61	1.35
16:1	0.27	0.46	1.03

Table 7-1 : Bandwidth v. Q-factor for first-order tuned circuit.

7.2.3 Initial specification

The specification of the sounder evolved, to some extent, as a result of certain decisions that were made whilst the hardware was being designed and tested. The hardware design will be discussed in chapter 9 and appendix A7. The specification currently adopted is as follows.

7.2.3.1 Transmitter

A wideband sequence is transmitted, consisting of 256 or 512 binary bits, clocked at 500kHz. It is based on a PRB sequence but there are no restrictions on its format. The transmitter PA is a half-H-bridge driver, and the antenna is an untuned air-cored induction loop.

7.2.3.2 Receiver

The receiver samples the input at 500kHz and collects a dataset of 2^{17} (131072) 16-bit samples, taking 262ms to do so. It has the capability to signal-average 128 datasets, with a total period of 34s. In theory, this will increase the SNR by $\sqrt{128}$ times, or 21dB.

A 131072-point Fourier transform will allow spectrum information to be presented at 3.8Hz intervals. (Although the spectral bandwidth will depend on the windowing function to be used).

7.3 Design Calculations

We need to ascertain the required size of transmitter and receiver antennas, and we need to obtain an estimate of the path attenuation and atmospheric noise contribution.

7.3.1 Antenna Design

In the previous two chapters we derived some design rules for transmitter and receiver antenna, which we can now apply to the design of a wideband sounder. Because of the extreme wideband nature of the system we will need to operate the transmitter antennas untuned and the receiver antenna either untuned, or tuned and damped.

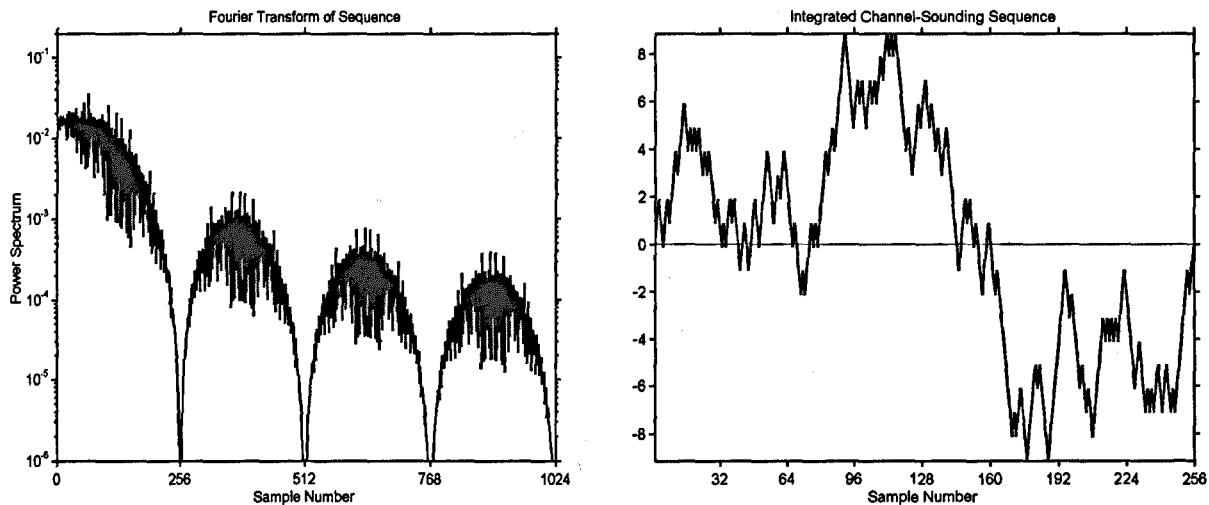
Clearly the reason for extensively analysing the transmitter and receiver antenna was to allow us to produce a design for a workable channel-sounder. However, as we shall note later, the project is not complete and there are a number of design issues still not decided, so it would be premature to present antenna designs at this stage. This is clearly a topic for future work.

7.3.2 Choice of Sounding Sequence

A 256-bit sequence is listed and discussed in chapter 8. However, as noted above, we now plan to use a 512-bit sequence for preliminary tests. The Fourier transform of the periodic 256-bit sequence is shown in *Figure 7-2a* below. We can observe that, although it is not a PRB sequence it has a similar power spectrum, of the form

$$\left(\frac{\sin n\pi/M}{n\pi/M} \right)^2 \quad (7-2)$$

where n is an integer and M is the length of the sequence.



*Figure 7-2 : Spectrum and antenna current due to 256-bit sounding sequence.
(Note: the spectrum is that of the periodic 256-bit sounding sequence and not that of the antenna current, which is the integrated sequence).*

Because we will be driving an untuned antenna, which we expect to be highly inductive, the current in the antenna will be the integral of the sequence. This is shown in *Figure 7-2b*. The obvious observation is that the runs of consecutive ones and zeros in the sequence result in the current building up in the antenna.

If we were to transmit a sequence 10101010..., such that the driving voltage swung between -1 and $+1$ units, then the current would ramp from $-1/2$ to $+1/2$ units, and the mean power would be $1/12$ unit. From a study of the waveform above, we can deduce that for the 256-bit sequence in question, the current ramps through a range of 18 units (not quite exactly ± 9) and the power is almost exactly 256 times that of the high-frequency square wave.

Although we will not present a proof here, it can be shown that the figure of 256 times the power is not a coincidence, and arises as a result of the nature of the sequence, which clearly contains energy over a broader spectrum than would a simple square wave. The increase in peak current of 18 times is important too, and dictates a modification to the design equations presented in chapter 5.

One of the conditions in the earlier analysis in chapter 5 was $Q \gg 1$. Clearly this is now ambiguous because we are no longer operating at a single frequency. If we define ΔT as the time duration of one ‘chip’ of the sequence then we can write

$$f_0 = \frac{1}{2\Delta T} \quad (7-3)$$

and use f_0 – the maximum square wave frequency – in the design calculations. If the sequence has length M then the maximum run-length of ones and zeroes is $\log_2 M$ and so the ‘high-Q’ approximation is equivalent to stating that

$$\frac{L}{R} \gg \log_2 M \cdot \Delta T \quad (7-4)$$

or, if the Q -factor is calculated at f_0 , this is

$$Q_0 \gg \pi \log_2 M . \quad (7-5)$$

Since our sequence will be 256 or 512 bits long, we require $Q_0 \gg 28$. However, this supposes that it is a requirement for the current to ramp linearly in the antenna. Clearly this need not be the case and, to a large degree, it does not matter what is the precise nature of the current waveform. We could run into synchronisation problems if the receiver antenna was not ‘matched’ to the transmitter in the sense that it differentiated the magnetic flux that arose from the integration of the sounding waveform. But, as we shall show in the next chapter, the system identification and synchronisation amounts to a ‘matched filter’ problem and we can suppose that it is possible to incorporate the antenna transfer functions into the matched filter calculation.

7.4 Integrating Sounding Into a System

The channel sounder is seen as an essential element of any adaptive, sub-surface, digital communications system. In this context, the primary physical parameters of the channel that normally affect the nature of the received signal are the attenuation and co-channel noise/interference.

We are expecting the signal attenuation to vary with frequency, depth and ground conductivity in the manner shown chapter 2. In addition, modelling suggests that the up-link signal propagation differs to the down-link. Noise and interference will also be a function of frequency, depth and conductivity but, being far-field signals, they are likely to be attenuated in a simple exponential manner, unlike the signal, which undergoes a complicated attenuation. It is not anticipated that multi-path propagation time dispersion will have a significant effect on the received signal, given the length of the paths considered and the likely data rates.

The proposed system adaptivity would take the following forms.

- i) Choosing an optimum operating frequency according to conditions (interference and sub-surface depth)
- ii) Utilising a different frequency for the up-link and down-link, due to the asymmetrical nature of the path.

- iii) Modification of the transmitted signal spectrum, via changes in transmitted data rate and/or modulation format, to overcome the effects of non Gaussian co-channel noise/interference.
- iv) Modification of the transmitted signal phase if we are operating our proposed method of interference cancellation, using 'quadrature phase space' as outlined in §6.5.
- v) Modification of the transmitted carrier frequency if we wish to dynamically adjust to the resonant frequency of the antenna, as its tuning drifts.
- vi) Variation in transmitted power level.
- vii) Variation in receiver antenna characteristics e.g. null-steering onto a dominant source of interference. (Cancellation of interference using an E-field antenna or a field gradient technique is also a possibility).
- viii) The possible use of pre-processing at the receiver to reduce the effects of non-Gaussian noise/interference.

In the first instance it is planned that an adaptive narrow-band transmission strategy based on i) above will be employed. Such a strategy will use single sideband (SSB) modulation, a narrow-band antenna and some form of 'hopping' within a limited set of alternative narrow-band channels.

Unfortunately, there are problems with such a system. Not least because the present generation of induction cave radios, at 87kHz, are operating in one of the few quiet portions of the spectrum. It is known, however, that interference varies greatly with geographical position; and experimenters in other parts of the UK, and in Europe, have discovered other frequency 'windows'. A second problem is that an induction loop antenna, tuned to resonance, does not easily lend itself to multiple frequency operation and so a method of driving an untuned antenna (which is inefficient) or of automatically tuning the antenna (which is cumbersome) will have to be investigated.

For these reasons, the use of spread-spectrum transmission will also be considered; in this case, the reduced gain of the required wideband antenna will be compensated by the processing gain of the spread-spectrum technique. Thus, a relatively inexpensive DSP technology will be used to compensate for relatively expensive and difficult to achieve r.f. performance. The system concept assumes that the channel sounding and communications processes will make use of common signal generation, processing and r.f. sub-systems, thus reducing overall system cost and complexity.

7.5 Concluding Remarks

We have outlined the design of a wideband channel sounder; and we envisage that embedded channel sounding is a key feature of an adaptive communications system for sub-surface operation.

The next phase of the research programme will be to take the prototype system, still under construction, and to use it to validate the propagation model via practical sounding measurements. A complete adaptive data transmission system architecture will then be tested.

The system design philosophy presented here owes much to designs for adaptive systems operating over other highly dispersive and time-variable paths, e.g. HF (2 – 30MHz) ionospheric skywave paths. However, in the sub-surface context, multi-path effects are not expected to be a significant problem; successful communication is likely to depend primarily on correct choice of frequency and techniques for mitigation of high levels of non-Gaussian noise and interference.

Before describing some preliminary results obtained with the prototype channel sounding receiver we will discuss, in detail, the sounding sequence and its use for system identification and synchronisation.

7.6 References

- Laflin, M. G. (1992). *ELF/VLF Spectrum Measurements*. Conference Proceedings 529 on ELF/VLF/LF Radio Propagation and Systems Aspects: Advisory Group for Aerospace Research and Development; NATO.
- Wellstead, P. E. (1975). Pseudonoise Test Signals and the Fast Fourier Transform. *IEE Electronics Letters* 11(10).

8 Sequence Design for System Identification

Abstract: *System identification using a periodic pseudorandom sequence is well-established. The procedure relies on the perfect (impulsive) nature of the auto-correlation function of the test sequence. It is possible to use a non-ideal test sequence provided that the input-output cross-correlation is carried out between the applied sequence and its inverse reference. A simple algorithm is presented for calculating the inverse of an arbitrary sequence under the operation of cross-correlation.*

We have described a channel sounder that transmits a wideband signal and uses signal-averaging at the receiver. To take advantage of the signal averaging it is important that the receiver maintains accurate synchronism with the transmitter. Synchronising to an external signal (e.g. a broadcast radio station) would be possible, but it was decided to utilise a delay-locked, or code-locked loop (CLL) for this purpose. It was therefore necessary to overcome the problem that the sounding sequence has a non-ideal (non-impulsive) auto-correlation function (ACF).

In this chapter we will discuss the use of non-ideal sequences for system identification. We will develop a simple algorithm for finding the inverse of a sequence, and will look at the noise factor and noise-equivalent bandwidth of systems using such sequences.

8.1 Inverse Sequences

An accurate estimate of the impulse response of an unknown linear time-invariant system may be made by applying a periodic pseudo-random test sequence and cross-correlating the system output with a replica of the test sequence (*Figure 8-1*). This well-established technique relies for its operation on the impulsive nature of the auto-correlation function (ACF) of the test sequence (Fan and Darnell, 1996; Godfrey, 1993)

It has been shown that it is also possible to obtain good results using a periodic sequence with imperfect ACF by following the output cross-correlator with a convolver that applies a correction signal (Al-Dabbagh, Darnell, Noble and Farquhar, 1997). We will demonstrate, below, that these two operations can be combined into a single correlation and, at the same time, that the mathematical analysis can be simplified.

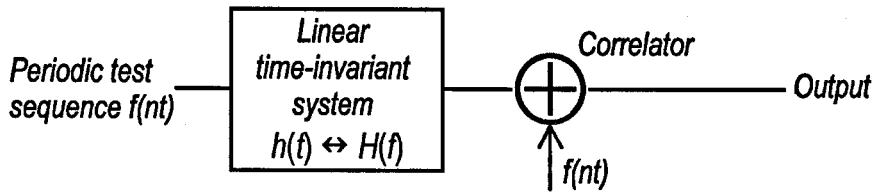


Figure 8-1 : System configuration for identification using test sequence.
The test signal $f(nT)$ is assumed to have perfect auto-correlation

8.1.1 Derivation of an Inverse Sequence

The test signal applied to the system is a periodic sequence of N samples with sampling interval T , represented by $f(nT)$. To obtain the system impulse response we need to cross-correlate the output signal with a time sequence that is an ‘inverse’ of f in the sense that

$$f(nT) \oplus f^{-1}(nT) = \delta(nT) \quad (8-1)$$

where \oplus denotes correlation and δ is unity at $nT=0$ and zero otherwise.

This equation can be solved with the aid of the Fourier transform and the convolution theorem. Later analysis will require us to apply Parseval’s theorem. The use of these operations on a discrete periodic function can lead to confusion because there is more than one way of defining them. For example, we choose to use ‘normalised’ versions of correlation and convolution. The relevant definitions are given in appendix A4.1.

Solving (8-1) for the inverse of f is straightforward. First, we apply the convolution theorem and re-arrange to give

$$F^{-1}(m\Omega) = \frac{N}{\tilde{F}(m\Omega)} \quad (8-2)$$

where \tilde{F} denotes a reversal of the order of the coefficients of F due to the fact that the operation in (8-1) is correlation and not convolution. We note that because f contains only real coefficients, reversing the order of the coefficients of F is equivalent to using the complex conjugate, F^* (appendix A4.2). Using this, and taking the inverse DFT of both sides results in

$$f^{-1}(nT) = IDFT\left(\frac{N}{F^*(m\Omega)}\right) \quad (8-3)$$

8.1.2 Verification of Result

The 256-bit sequence shown in **Table 8-1** has the imperfect periodic ACF of **Figure 8-2** with sidelobes up to -21 dB relative to the correlation peak. The r.m.s. value of the sidelobe function, considered separately to the peak, is -30 dB relative to the correlation peak and the

r.m.s. of the peak function is obviously -24dB ($1/\sqrt{256}$), thus the peak / sidelobe energy ratio is 6dB . Equation (8-3) was used to calculate the inverse using a MatLab program, the salient line of which was

```
inverseFn = ifft( N / conj(fft(test_signal)) )
```

where 'test_signal' and 'inverseFn' are vector quantities.

Using Matlab's long-floating-point format, the correlation of this inverse with the original sequence showed the expected ideal response, with the correlation peak at unity and the sidelobes all below 10^{-16} . The inverse is shown in *Table 8-2* where it has been rounded to two decimal places. With these rounded coefficients, the sidelobes of the correlation are still below -69dB .

1	0	1	1	1	0	0	0	1
1	0	0	1	1	1	0	0	0
1	1	1	0	1	0	1	0	1
0	1	0	1	0	0	1	1	1
0	1	0	1	0	1	1	1	0
1	0	0	1	0	1	1	1	0
0	1	1	0	0	1	0	0	1
1	0	0	0	1	0	0	0	1
1	0	0	0	0	1	0	0	0
1	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	1

*Table 8-1 : 256-bit test sequence $f(nT)$.
Values read across then down*

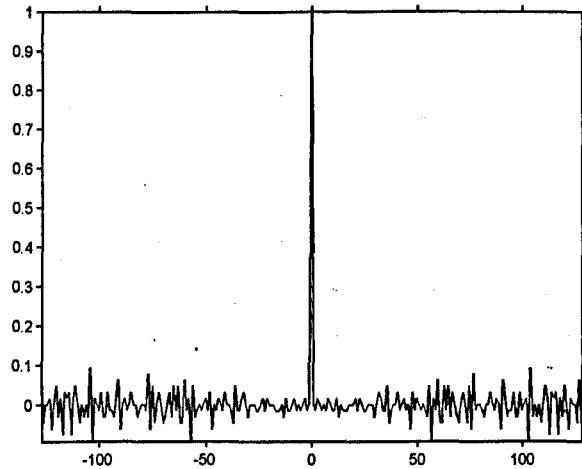


Figure 8-2 : Auto-correlation of a bipolar version of the sequence listed in Table 8-1.

-0.59	-0.64	+2.18	+1.55	-2.61	-2.11	-1.00	+0.04	+3.62	+0.53
+2.58	-2.03	+2.69	-1.79	-1.83	-1.13	-3.08	+1.01	+1.40	+2.82
+1.05	+1.88	+2.90	+1.57	+3.13	-0.96	-2.25	+1.79	-2.31	-1.31
-2.10	-2.75	+1.69	-2.02	+0.42	+2.05	-4.10	+0.39	+2.21	+2.47
+1.91	+1.47	+1.71	-3.07	+2.53	-2.25	+0.76	-0.61	+0.89	-0.80
+1.65	+3.48	-0.13	-1.82	-1.49	-4.20	-1.75	-0.17	+1.57	+4.11
-5.48	-1.04	-3.00	+4.39	-3.07	+3.34	-3.43	+3.93	+0.70	-1.70
-1.63	+1.67	+3.06	-5.11	0	+1.59	-0.57	+1.49	+2.52	+1.50
+2.39	+1.75	+3.45	-4.81	+3.86	+1.23	+2.77	+0.87	-2.65	-1.46
+3.06	+1.46	-1.92	+2.35	+2.18	+2.43	-2.29	+1.42	+2.10	+1.09
-3.08	+0.28	+0.22	-1.43	+1.05	-1.13	+3.54	-4.84	-0.31	+1.17
-2.62	+3.65	-2.97	-3.08	-0.71	+2.05	-1.80	-2.91	+2.37	-3.19
+5.01	+0.85	-1.12	-0.50	-1.53	-1.05	-3.55	+4.00	+1.04	-1.96
-3.00	+3.36	+1.25	+2.72	-3.82	-1.49	+2.91	+1.91	+2.06	+1.54
-2.54	+1.49	-5.22	+3.93	-0.44	+0.35	+1.02	+1.98	-1.14	-2.32
-2.76	-0.24	+1.21	-1.44	-3.83	-0.94	+2.13	-1.94	+2.25	+1.00
+3.04	-2.09	+1.95	-1.59	+0.32	+3.00	+1.87	+1.77	-1.24	+0.01
+3.66	-4.73	+5.57	-1.30	+4.17	+2.10	+2.45	-1.73	-2.22	-4.35
+0.63	+0.44	+3.70	-4.67	+4.16	-0.59	-4.27	+3.63	+1.01	-0.19
+0.70	-0.53	-1.57	+4.47	-2.20	+1.96	+1.26	-0.91	+2.42	-3.06
+1.81	-1.56	-0.99	-4.01	-0.65	-3.38	+2.44	+0.12	-4.09	+3.10
+0.75	+1.97	-0.53	+0.03	+3.90	-2.81	-3.61	+3.07	-2.07	-0.56
+0.01	+1.52	-3.62	+1.15	+1.88	-1.41	-2.18	-1.11	-1.09	-1.56
-3.55	+3.00	-0.66	+4.12	-0.35	-1.29	-0.86	-2.22	+2.17	-1.82
-2.17	-0.16	-0.63	+2.22	+2.48	-1.36	+0.53	-7.71	+5.26	-2.46
-1.25	-3.23	-2.12	-0.75	-1.78	-2.08.				

Table 8-2 : Inverse sequence, rounded to 2 decimal places.
Values read across then down

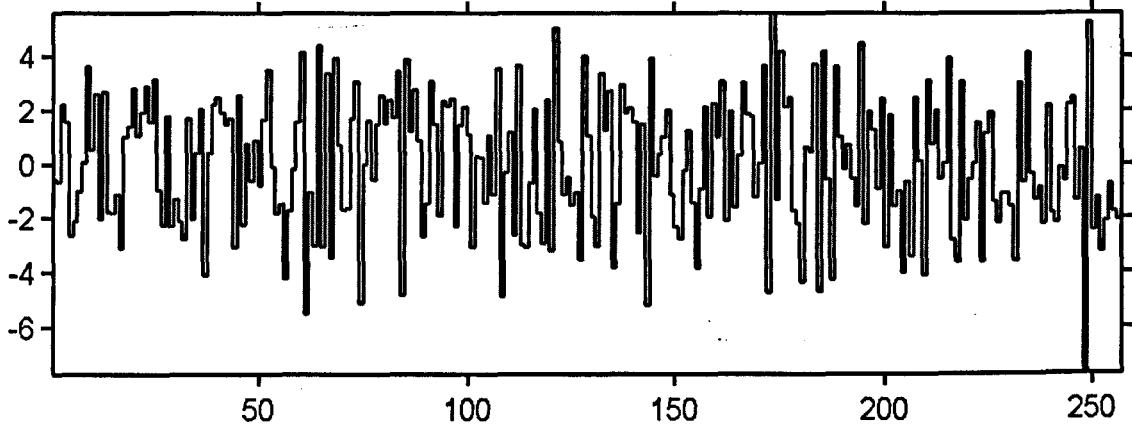


Figure 8-3 : Plot of the inverse sequence listed in Table 8-2.

8.1.3 Observations

8.1.3.1 On the sum of coefficients

The sum of the coefficients in **Table 8-2** is exactly 2; this arises because the product of the sum of the coefficients of f and its inverse, respectively, is N . This statement is proved by noting that the mean of the coefficients of f , given by $1 \oplus f$, represents a d.c. level; the DFT definition we are using equates this to $F(0)/N$ so the sum of the coefficients of f is $F(0)$ and thus we have a restatement of (8-2) for the special case of $m\Omega=0$.

8.1.3.2 On the existence of an inverse

A second observation, from (8-3), is that the inverse does not exist if $F(m\Omega)$ contains one or more zeroes. One case where F contains a zero is where the test signal is required to have no d.c. component, e.g. in line transmission systems. This problem was mentioned in (Al-Dabbagh *et al.*, 1997) without elaboration. However the situation can be resolved – we use $f(nT)$ as the test signal, which we use to calculate the inverse function, but we note that the d.c. component will be removed, so the correlation is essentially

$$\left\{ f(nT) - \frac{F(0)}{N} \right\} \oplus f^{-1}(nT) = \delta - \frac{F(0)}{N} \oplus f^{-1}(nT) \quad (8-4)$$

and thus the effect is simply to add a constant to the output.

8.1.3.3 When the mean of f is zero

From §8.1.3.1 we can see that the right-hand side of (8-4) is simply

$$\delta - 1/N \quad (8-5)$$

which also has a mean of zero. This shifted delta function is the cross-correlation product when we use a copy of f that has been level-shifted so that the mean of its coefficients is zero.

It is interesting to look at the situation where, in addition to removing the d.c. component of f , we shift the d.c. level of the inverse function by an arbitrary amount, k . Now an extra term appears on the right of (8-4), of

$$k \oplus \left\{ f(nT) - \frac{F(0)}{N} \right\} \quad (8-6)$$

but this is proportional to the mean¹ of the expression in brackets, which is zero by definition. Hence we can deduce that, with no d.c. component to f , the d.c. level of the inverse f^{-1} does not affect the correlation product. We can turn this around and say that if the d.c. component of the inverse is removed then the d.c. level of the incoming sequence does not matter, and we have a precisely defined offset to the signal. This observation, although interesting, is not of significance in our application, where the system identification is being used in conjunction with a code-locked loop (CLL) that maintains synchronisation. CLLs utilise two correlations, taking the difference between an ‘early’ and a ‘late’ version. The constant term in (8-5) is therefore eliminated, whether it is precisely defined or not.

Equation (8-5) reminds us that the (normalised) ACF of a bipolar M -sequence (values ± 1) has a correlation peak of unity and sidelobes of $-1/M$, and that we can write it as

¹ Because, as observed earlier, a constant, correlated against sequence f is proportional to the mean of f .

$$\frac{M+1}{M}\delta - \frac{1}{M} \quad (8-7)$$

Working backwards from this, using (8-4), we can see that the CCF of a bipolar (± 1) and a unipolar ($0, +2$) M -sequence is $\delta \cdot (M+1)/M$, which only needs a scale-factor adding, to allow us to say that a unipolar maximal-length binary sequence of length M has an inverse that is a bipolar copy of the same sequence; a result derived in more general terms, in (Al-Dabbagh, O'Farrell and Darnell, 1998).

8.2 Noise Factor

The operation of cross-correlation to produce a system identification signal is analogous to the operation of a matched filter. We should therefore expect the noise factor of the system to be degraded if our sequence f is not self-inverse. This will be confirmed in the following analysis, for which we will adopt a more general definition of the correlation, namely

$$f(nT) \oplus h(nT) = r(nT) \quad (8-8)$$

For system identification, r will be the delta function, but for the operation of a code-locked loop we may require r to be some other sequence, as will be discussed later.

8.2.1 Derivation of Signal/Noise Ratio

At the output of the correlator, we ‘integrate and dump’. The signal is therefore proportional to $r(n_i T)$ where the value of n_i represents the phase shift of the h sequence with respect to f . For convenience later, we can write

$$r(n_i T) = [f(nT) \oplus h(nT)]_i \quad (8-9)$$

The mean energy-per-bit in the signal f is found by calculating the mean square of f ,

$$E = \frac{T}{N} \sum_{n=0}^{N-1} |f(nT)|^2 \quad (8-10)$$

Sampled Gaussian noise with a spectral density of n_0 [W/Hz] is applied to the receiver correlator resulting in an output power spectral density² [W/Hz] of

$$n_0 \cdot \frac{|H(m\Omega)|^2}{N^2} \quad (8-11)$$

The spectral density of the input noise was based on a bandwidth of $1/2T$, so the total noise power, at the output of the system, is

2 The reason for the factor of N^2 in the following equation is discussed in appendix A4.4.

$$n = \frac{1}{2} n_0 \frac{1}{N^2 T} \sum_{m=0}^{N-1} |H(m\Omega)|^2 \quad (8-12)$$

We need to express this in terms of $h(nT)$ which we can do by invoking Parseval's theorem (see appendix A4.3). to obtain

$$n = \frac{1}{2} n_0 \frac{1}{NT} \sum_{n=0}^{N-1} |h(nT)|^2 \quad (8-13)$$

We can combine the above expressions to allow us to write the signal/noise power ratio as

$$\frac{s^2}{n} = \frac{2E}{n_0} \frac{r(n_i T)^2}{\frac{1}{N} \sum_{n=0}^{N-1} |f(nT)|^2 \cdot \frac{1}{N} \sum_{n=0}^{N-1} |h(nT)|^2} \quad (8-14)$$

8.2.2 Analogy to Matched Filters

The last term in (8-14) is in the time domain but it is interesting to look at a frequency-domain representation. Using the convolution theorem on (8-8) we obtain

$$\tilde{r} = IDFT \left(\frac{1}{N} F^*(m\Omega) H(m\Omega) \right) \quad (8-15)$$

where F^* is the complex conjugate of F . We can expand this to

$$\tilde{r}(n_i T) = \frac{1}{N^2} \sum_{m=0}^{N-1} F^*(m\Omega) H(m\Omega) e^{jmn_i \Omega T} \quad (8-16)$$

where the $1/N^2$ arises as a result of the definition of the DFT pair. The bottom terms in (8-14) can be treated using Parseval's theorem, whereupon each gains a $1/N$ term to cancel the above N s. The resulting expression is

$$\frac{\left| \sum_{m=0}^{N-1} F^*(m\Omega) H(m\Omega) e^{jmn_i \Omega T} \right|^2}{\frac{1}{N} \sum_{m=0}^{N-1} |F(m\Omega)|^2 \cdot \frac{1}{N} \sum_{m=0}^{N-1} |H(m\Omega)|^2} \quad (8-17)$$

which is similar to that which arises in the derivation of a matched filter; see, for example, (Schwartz, 1980) §5.6. Quoting Schwarz's inequality and noting, in this instance, that the complex conjugate of F already appears on the top, this expression reaches a maximum when

$$H(m\Omega) = (F^*)^*(m\Omega) e^{jmn_i \Omega T} \quad (8-18)$$

In our case, we are interested in the signal at $n_i = 0$, (i.e. zero shift in the correlator) so the optimum SNR is when $H = F$.

The problem is usually presented this way because we desire to find the value of H that maximises the expression for a given F and, from this, demonstrate the concept of a matched

filter. However, in our case, we already know H and it is simpler to evaluate the SNR using the time-domain expression in (8-14), namely

$$\frac{r(n_0 T)^2}{\frac{1}{N} \sum_{n=0}^{N-1} |f(nT)|^2 \cdot \frac{1}{N} \sum_{n=0}^{N-1} |h(nT)|^2} \quad (8-19)$$

which we can term the ‘noise factor’ of the system, since it has a maximum value of unity and expresses the degradation in SNR caused by using an unmatched system. The condition for maximum noise factor is now $h = f$.

8.2.3 Investigation of Noise Factor Expression

It is straightforward to evaluate the noise factor for the case of a maximal-length M -sequence cross-correlated against itself, and to show that in this ‘matched’ condition the noise factor is unity, or 0dB.

Such a sequence is not necessarily self-inverse. For example, a bipolar M -sequence with coefficients of ± 1 has an ACF of

$$f_{M,\text{bipolar}} \oplus f_{M,\text{bipolar}} = \frac{M+1}{M} \delta - \frac{1}{M}. \quad (8-20)$$

However, a 255-bit M -sequence with coefficients of $+0.9315\dots$ and $-1.0645\dots$ is self-inverse and an evaluation in MatLab shows that the noise factor is unity, as expected. (The coefficients can be evaluated as $(-1 \pm 15) \times (\sqrt{255})/(16 \cdot 15)$, and we are using a normalised correlation.)

Even if f is an M -sequence, Schwarz’s inequality tells us, as noted above, that the noise factor can be less than unity if f not self-inverse. As an example, consider the bipolar M -sequence f , as above. Cross-correlating a scaled copy of this with a unipolar version of the sequence (coefficients 0 and 2), thus...

$$\begin{aligned} & \frac{M}{M+1} f \oplus (f+1) \\ &= \delta - \frac{1}{M+1} + \frac{M}{M+1} (1 \oplus f) \\ &= \delta \end{aligned} \quad (8-21)$$

shows that the two form an inverse pair³. Evaluating the noise factor for this pair of sequences shows that it is $128/255$, or -3.0 dB. This can be attributed to the fact that the zeroes in the unipolar sequence do not contain any signal energy.

In summary, considering just M -sequences, we have three possibilities...

- i) The sequence is self-inverse; using $f \otimes f$ leads to unity noise factor and a perfect ACF.
- ii) The sequence is not self-inverse; using $f \otimes f$ leads to unity noise factor, but the ACF has a constant level superimposed on it.
- iii) The sequence has an inverse h ; using $f \otimes h$ leads to a perfect ACF, but a less-than-ideal noise factor.

8.2.4 Sequences with imperfect ACF

Central to this analysis is the question of what happens to the noise factor if f is not a self-inverse M -sequence. Consider a unipolar M -sequence for $M = 255$, with one bit added (Table 8-1). The resulting ' $M+1$ ' sequence has an imperfect ACF (Figure 8-2) and a calculable inverse, h (Table 8-2). Cross-correlating these two sequences produces a noise factor of 0.33, or -4.8dB which, as in the example of §8.2.3, is due to the zero coefficients in f . However, it is often the case with communications systems that we must use a test sequence that contains no d.c. component. As noted previously, this would render an inverse uncalculable; but we can proceed by using the inverse of the unipolar sequence, and substituting a bipolar version of the test sequence f , constructed by subtracting 0.5 from the coefficients. We now obtain a noise factor of 0.656 or -1.83dB , and the expected cross-correlation product of $\delta - 1/256$. We can further maximise the noise factor by a small amount by using a copy of f^{-1} with no d.c. component. It was noted in §8.1.3.3 that if the d.c. level of f were zero then the d.c. level of f^{-1} would not affect the cross-correlation product; however, the noise factor is improved by this action. The reason for this is that for any set of integers a , the sum $\sum(a - m)^2$ is a minimum when the constant m is the mean of the set, i.e. $m = \sum a/N$ or equivalently $\sum(a - m) = 0$. In this particular case the improvement in noise factor is minimal because f^{-1} has only a small d.c. level to begin with.

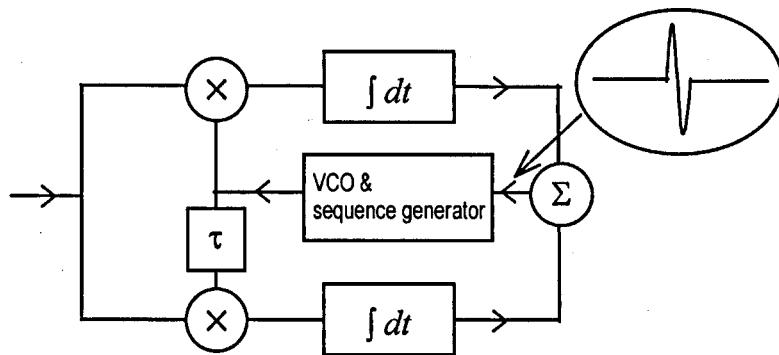
8.3 System Synchronisation

8.3.1 Code-Locked Loop

We have seen that system identification can be obtained by cross-correlating the test sequence with its inverse but that in certain circumstances a d.c. level is imposed on the otherwise perfectly impulsive identification signal. The problem of system synchronisation is similar to that of identification but a simple correlation is not sufficient because it is not easily possible to

3 We make use of the fact, noted in §8.1.3.1, that $1 \oplus f$ is the mean of the coefficients of f , which is $1/M$ for the sequence f in question. This result was also discussed at the end of §8.1.3.3.

detect the polarity of the phase error at the receiver. A traditional solution is to employ an additional correlator with a time delay of one or, sometimes, two chips of the sequence. (The binary digits comprising a PRB sequence are referred to as chips.) The output from the pair of correlators is differenced (*Figure 8-4*), resulting in a bipolar synchronisation control signal.



*Figure 8-4 : A traditional code-locked loop.
This employs a pair of correlators that give rise to a bipolar control signal with a large dead-band.
The delay τ is one or two 'chips'.*

Being a bipolar signal, this can be used as the control input for a variation on the phase-locked loop, known as a delay-locked or code-locked loop (CLL). Synchronisation occurs when the two correlations are at $\pm \frac{1}{2}T$ where T is the symbol interval of the sequence; hence the correlations are referred to as 'early' and 'late' relative to the transmitted sequence. If the receiver is more than $1.5T$ out of phase with the transmitter then the output of the code-detector will be governed only by noise, one source of which is due to channel imperfections, thus the acquisition time of the loop could be high.

Acquisition time is clearly more of a problem at low frequencies. A GPS receiver correlates a 1023-bit sequence, clocked at 1.023MHz, in 1ms and can easily obtain lock in a few seconds. However a similar system with a clock of 1kHz would take over 17 minutes to cycle through all shifts of the sequence.

8.3.2 An improved code detector

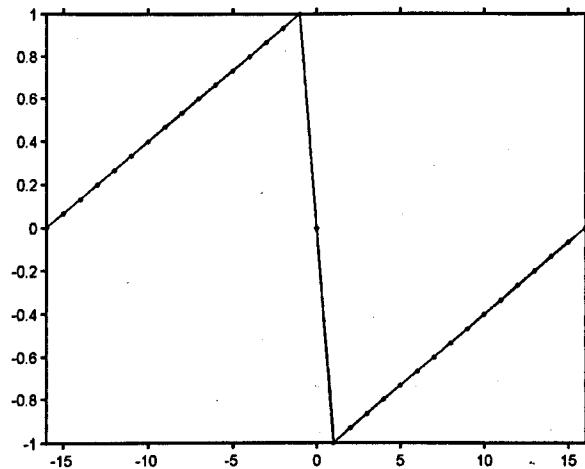
To improve the acquisition time of the code-locked loop, we need to be able to measure the transmitter/receiver phase error outside the $1.5T$ or $2T$ band in which the traditional code-detector works. Various fast-acquisition schemes have been devised, but an attractive solution must be to choose a pair of sequences that, when cross-correlated, give rise to an improved system control signal. For example, the sequence $r(nT)$ defined by

$$\begin{aligned}
 r(nT) &= \frac{n + \frac{1}{2}N}{\frac{1}{2}N - 1}, \quad -\frac{1}{2}N \leq n < 0 \\
 r(0) &= 0 \\
 r(nT) &= \frac{n - \frac{1}{2}N}{\frac{1}{2}N - 1}, \quad 0 < n \leq \frac{1}{2}N
 \end{aligned} \tag{8-22}$$

and shown in *Figure 8-5* below has a similar characteristic, interpolated over the range $-1 < n < 1$ to a conventional early-late correlator pair with a difference in delay of 2 symbol periods, but no deadband. (The reason for choosing a delay of 2 chips is that it simplifies some of the following analysis). The CLL now takes on the form of *Figure 8-6* below. The control signal $r(nT) \leftrightarrow R(m\Omega)$ is as defined in (8-8) and it is straightforward to show that the function we need to correlate against f to get r is, in similar fashion to (8-3)

$$h(nT) = N \cdot \text{IDFT} \left(\frac{R * (m\Omega)}{F * (m\Omega)} \right). \tag{8-23}$$

This result provides a straightforward means of calculating the required receiver sequence to give any desired control signal. Unlike a conventional CLL, this system does not restrict us to using a PRB sequence for f . We can use an arbitrary sequence, provided that it contains no zeroes in the frequency domain that would render (8-23) infinite.



*Figure 8-5 : A code-locked loop control signal for with no dead-band.
There is a potential improvement in acquisition time. The in-lock condition is at n=0,
and the code (in this example) is 32 chips in length.*

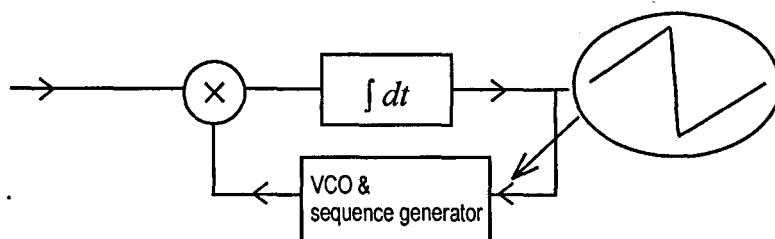


Figure 8-6 : Code-locked loop with single multi-valued cross-correlation.

8.3.3 Ripple caused by zeroes in R

From (8-23) we can see that if there is a zero in F (at $m = k$, say) then the equation might not be solvable. However, in the frequency domain, (8-23) becomes

$$F^*(m\Omega) \cdot H(m\Omega) = N \cdot R^*(m\Omega) \quad (8-24)$$

showing that if there is a corresponding zero at $R(k\Omega)$ we *can* solve the equation, assigning an arbitrary value to $H(k)$. Conveniently, we can define $R(k\Omega)/F(k\Omega) = 0$ if $R(k\Omega)=0$, regardless of $F(k\Omega)$.

Conversely, if there is a zero at $R(k\Omega)$ then the value of the corresponding coefficient in F is immaterial. The particular significance of this is that if r contains no d.c. component (so $R(0) = 0$) then the d.c. level of the transmitted sequence is not important. We could have arrived at this conclusion by noting that the ACF of a PRB sequence contains a d.c. level, but that this cancels when an ‘early-minus-late’ correlation is performed.

If a zero at $F(k\Omega)$ is not matched by a zero in R then the equation is not solvable unless we alter R , re-defining $R(k\Omega) = 0$ and thereby introducing a ripple into the output of the code detector. From the particular definition of the DFT we are using, we can see that removing a frequency component $\delta((m-k)\Omega)$ from $R(m\Omega)$ will result in a ripple in $r(nT)$ of $(1/N)\exp(j2\pi mn/N)$.

8.3.3.1 Example: Ripple in control signal

The above discussion is amply demonstrated in the following example. We begin by defining an 8-bit bipolar sequence and we use it to construct the 32-bit sequence $[n_1 n_1 n_1 n_1 n_2 n_2 n_2 \dots]$ (see **Table 8-3** below) which will have no component at $F(0)$, $F(8\Omega)$, $F(16\Omega)$ and $F(24\Omega)$ in the frequency domain. Suppose that we wish the output of the code-detector to be the function $r(nT)$ defined by (8-22) and Figure 8-5 above. We can observe that

- i) $R(0) = 0$ and $R(16\Omega)=0$ so the absence of a d.c. component in f (i.e. $F(0) = 0$) is not a problem. We simply define $R(0)/F(0) = 0$ in the calculation of h . A similar situation applies for $F(16\Omega)$.

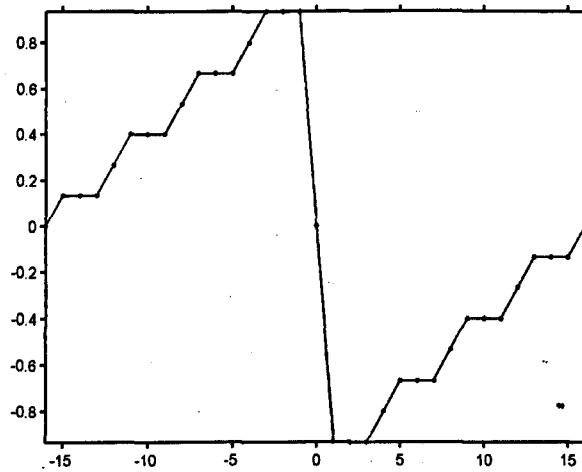
- ii) The presence of a frequency coefficient at $R(8\Omega)$ and the lack of the corresponding components in F means that $R(8\Omega)/F(8\Omega)$ is infinite and so h cannot be calculated. To remove this difficulty we redefine $R(8\Omega)$ to be zero and then define $R(8\Omega)/F(8\Omega) = 0$ as in (i). A similar situation applies for $F(24\Omega)$.

We can now calculate h (*Table 8-3* below). Performing the correlation $f \oplus h$ gives rise to the function shown in *Figure 8-7* below, which is similar to the desired function of *Figure 8-5* but with a ripple component as predicted from our analysis.

This is something of a contrived example because even if f is not a PRB sequence we can probably arrange for it to contain no zeros in the frequency domain, apart from $F(0)$; and, if $R(0) = 0$ as well, there is no problem.

a) Sequence generator = [-1 -1 -1 +1 +1 +1 -1 +1] is used to derive...
$f = [\begin{array}{cccccccccccccccc} -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 \end{array} \dots]$
b) Sequence $h = [\dots]$ $+4.8000 \quad 0 \quad 0 \quad +1.6000 \quad +1.6000 \quad 0 \quad 0 \quad -3.7333 \dots$ $-3.7333 \quad 0 \quad 0 \quad -2.6667 \quad -2.6667 \quad 0 \quad 0 \quad -3.7333 \dots$ $-3.7333 \quad 0 \quad 0 \quad +1.6000 \quad +1.6000 \quad 0 \quad 0 \quad +4.8000 \dots$ $+4.8000 \quad 0 \quad 0 \quad -2.6667 \quad -2.6667 \quad 0 \quad 0 \quad +4.8000]$

*Table 8-3 : Generator sequences for control signal in example in §8.3.3.1.
The correlation product of these two sequences is the control signal shown in Figure 8-7 below.*



*Figure 8-7 : Control signal for example in §8.3.3.1.
This control signal is the closest we can achieve to that of Figure 8-5 with the sequence of Table 8-3a above, because the zeroes present at $F(8\Omega)$ and $F(24\Omega)$ are not matched by zeroes in $R(m\Omega)$, the FFT of the signal in Figure 8-5.*

8.3.3.2 A zero dead-band early/late control signal

A 256-bit ' $M+1$ ' sequence and its inverse were described in §8.1.2. Using a bipolar version of that sequence, we can construct the 'inverses' required to give an early/late control signal, and the 'improved' control signal. These are shown in **Figure 8-8 / Table 8-4** and **Figure 8-9 / Table 8-5** respectively. Further comment will be made when we have discussed the noise performance of a code-locked loop.

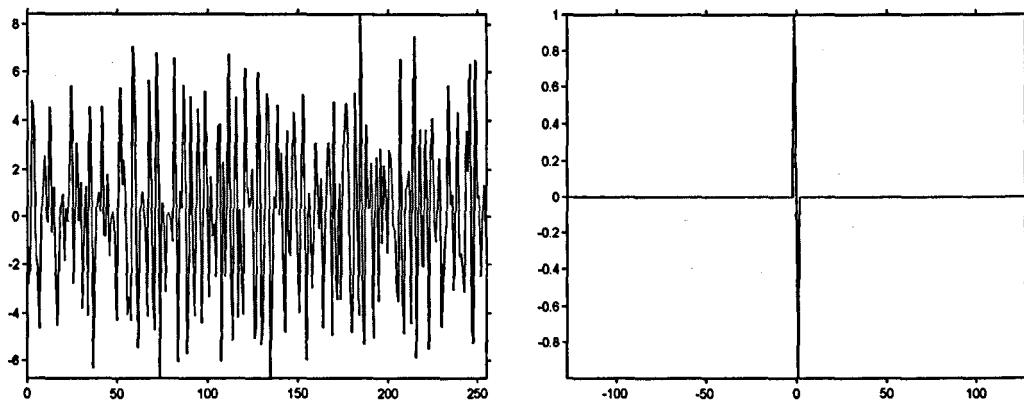


Figure 8-8 : Inverse sequence needed to achieve the conventional CLL 'S-curve' control signal.
Left: a plot of Table 8-4 below which, when correlated with a bipolar version of the $M+1$ sequence of Table 8-1 gives rise to **Right:** the 'early/late' control signal.

-1.43	-2.77	-2.19	+4.79	+3.65	-1.61	-2.15	-4.62	-0.49	+1.04	+2.56	-0.11
-0.23	+4.52	-0.66	+1.25	-2.14	-4.48	-1.82	+0.35	+0.94	-1.85	+0.31	-0.23
+2.53	+5.38	-2.75	+0.06	+3.10	-0.21	+1.43	-3.79	-0.73	+1.27	-4.07	+4.52
+1.66	-6.30	-2.08	+0.29	+1.01	+0.21	+4.54	-0.82	-0.83	+1.77	-1.64	-0.14
+0.19	-0.76	-4.28	+1.78	+5.30	+1.36	+2.38	+0.27	-4.03	-3.32	-4.28	+7.05
+5.15	-2.48	-5.43	+0.07	+1.05	+0.36	-0.58	-4.13	+5.63	+2.33	-3.38	-4.69
+6.78	+3.06	-6.70	+0.57	+0.10	-3.09	-0.01	+0.14	-0.24	-1.06	+6.56	-0.41
-6.05	+1.09	+0.36	+5.42	+2.33	-5.71	-2.92	+4.98	-0.89	-4.10	-0.08	+4.47
+1.01	-4.39	+0.34	+5.18	+0.81	-3.30	+1.72	-0.83	-0.31	-2.49	+3.72	+3.85
-6.01	+2.32	-2.48	+0.35	+6.72	-2.26	-5.12	+1.09	+4.96	-4.17	+0.27	-2.63
-4.04	+6.13	+1.35	+0.41	+0.55	+2.01	-5.05	-4.58	+5.95	+4.04	-5.31	-4.26
+0.64	+5.07	+4.21	-6.73	-3.40	+0.85	+0.37	+4.60	+0.06	+2.68	-2.45	-4.78
+3.58	-1.46	-1.63	+2.16	+4.30	+1.62	-2.08	-3.97	+1.20	+5.04	-0.50	-5.96
+1.00	-0.12	-2.94	-0.79	+3.09	+1.09	-0.50	+1.63	-4.59	-1.55	+1.23	+3.11
+1.76	-4.90	+4.74	-1.90	-3.42	+1.40	-3.41	+1.72	+3.83	+4.67	+2.62	-2.86
-4.80	-3.06	+5.12	-0.46	-4.08	+8.43	-4.22	-5.29	+3.82	+0.31	+0.34	+2.27
-5.00	+0.64	+2.51	-3.47	+2.87	-1.15	+2.15	+0.61	-1.51	+2.80	+2.46	-0.34
-0.63	-3.09	-3.50	+6.53	-2.98	-4.84	+1.13	+1.28	+1.94	-4.43	+2.84	+7.50
-5.88	-1.54	+3.64	-2.07	-2.08	+3.62	+0.37	-5.50	+2.55	+4.06	-0.30	-1.09
+0.45	+2.46	-4.55	-2.89	-1.13	-0.32	+5.42	+0.52	+0.92	-3.03	-0.40	+4.34
-1.66	-1.55	-2.38	-3.10	+3.58	+1.94	+6.34	-4.72	-5.25	+6.51	+0.78	+0.87
-2.48	-0.35	+1.33	-1.19								

Table 8-4 : data for Figure 8-8.
Note that the values $h(n)$ in this table – being those for an 'early minus late' correlation – can be obtained from the values $g(n)$ in Table 8-2, by the formula $h(n) = g(n-1) - g(n+1)$.

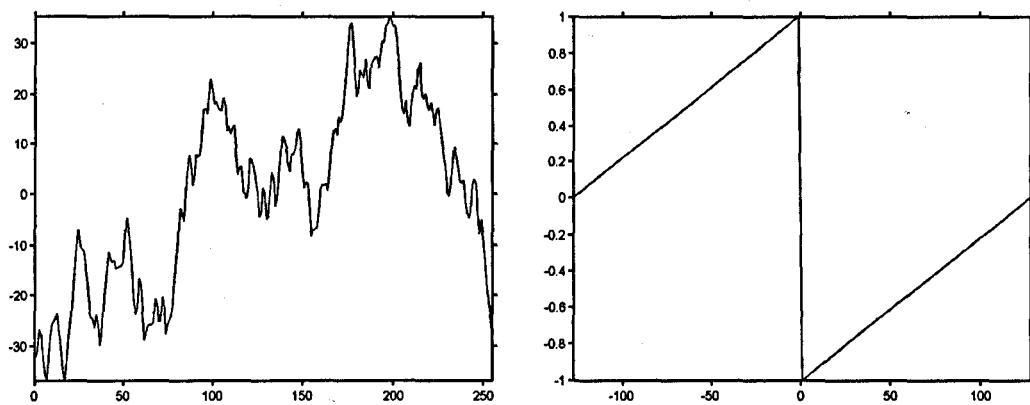


Figure 8-9 : Inverse sequence needed to achieve a zero dead-band control signal.
Left: a plot of Table 8-5 below which, when correlated with a bipolar version of the M+1 sequence of Table 8-1 gives rise to **Right:** the 'improved' control signal.

-30.72	-31.98	-30.44	-26.70	-27.78	-32.55	-35.70	-36.68	-33.00	-28.83	-25.71	-25.17
-24.51	-23.63	-27.29	-30.29	-34.55	-36.65	-34.24	-30.00	-26.11	-23.16	-18.35	-13.86
-09.14	-06.97	-10.23	-10.71	-11.25	-14.91	-18.36	-23.26	-24.34	-24.68	-26.30	-23.82
-25.90	-29.64	-27.04	-22.34	-17.94	-14.55	-11.36	-12.76	-13.32	-13.05	-14.57	-14.44
-14.16	-14.08	-13.23	-08.08	-04.71	-06.69	-10.04	-15.78	-21.79	-23.75	-22.36	-16.65
-18.05	-24.64	-28.73	-27.35	-26.03	-25.77	-25.88	-25.39	-20.75	-21.78	-25.16	-25.14
-20.38	-22.47	-27.63	-26.05	-25.03	-24.12	-20.08	-16.04	-12.14	-07.99	-02.77	-04.16
-05.14	-00.02	+04.00	+07.65	+05.84	+01.68	+03.28	+07.82	+07.34	+07.76	+12.32	+16.95
+17.08	+16.19	+19.73	+22.93	+20.91	+18.07	+18.56	+17.33	+16.92	+16.83	+19.25	+17.92
+12.72	+13.57	+12.09	+13.11	+13.77	+07.66	+03.83	+05.17	+05.40	+00.64	+00.09	-00.74
+01.08	+06.97	+06.68	+05.03	+02.97	+00.35	-04.29	-03.85	+01.21	+00.27	-04.74	-04.40
+00.23	+04.22	+03.09	-02.28	-00.87	+03.97	+07.96	+11.58	+10.55	+09.48	+05.70	+04.39
+07.89	+07.79	+09.15	+12.15	+12.98	+09.48	+04.34	+01.30	+02.27	+02.02	-03.30	-08.12
-06.94	-06.77	-06.47	-03.21	+00.85	+01.79	+01.64	+01.98	+00.69	+04.02	+08.91	+12.57
+13.09	+11.84	+15.53	+14.45	+15.28	+19.56	+22.43	+28.74	+33.31	+34.02	+30.02	+23.38
+19.62	+20.69	+24.84	+23.84	+23.30	+26.88	+21.96	+21.29	+25.95	+26.76	+27.26	+27.41
+25.28	+28.18	+30.45	+30.18	+33.42	+33.76	+35.26	+34.60	+33.32	+33.55	+30.97	+25.91
+21.19	+17.12	+16.15	+18.72	+14.70	+13.69	+17.56	+20.28	+21.72	+21.20	+25.14	+26.21
+19.73	+19.18	+20.18	+17.51	+16.93	+18.46	+16.33	+13.83	+16.87	+17.33	+13.70	+10.37
+08.14	+05.45	+00.29	-00.29	+02.05	+05.52	+09.32	+07.65	+05.47	+02.35	+02.28	+02.62
-01.42	-03.79	-04.59	-03.00	+01.72	+02.83	+01.98	-05.26	-07.75	-04.94	-08.69	-13.23
-18.64	-21.55	-24.11	-28.01								

Table 8-5 : data for Figure 8-9.

8.3.4 Noise factor of a code-locked loop

We saw in §8.2 how to express the noise factor of a 'matched sequence' system in a similar way to that of a matched filter. Unfortunately the expression for noise factor, (8-19), cannot be used with a code-locked loop because it depends on expressing a signal to noise ratio. When a code-locked loop is in lock there is no signal present at its output – that is $r(0) = 0$. The same is true of a conventional phase-locked loop⁴.

A PLL or CLL has a similar transfer function to that of a first order low-pass filter, and so the noise performance is characterised by the extent to which incoming noise is eliminated by the low-pass filter effect of the loop. The noise analysis of a phase locked loop (Blanchard,

4 Apart from a type 0 loop where it is only 'very close' to zero, depending on the system gain.

1976; Wolaver, 1991) and of a code-locked loop (Spilker, 1963) has been well studied, and it will not be covered in detail here. Instead, we can note that the noise-equivalent bandwidth (NEB) of the loop must depend, in some way, on the noise energy that passes through the correlator, which was given by (8-13). It will also depend on the gain of the phase-discriminator or correlator which we can write as dr/dn or $\frac{1}{2}(r(1) - r(-1))$. In all the examples given here, this factor is the same, so it will be ignored in this analysis.

We can scale the sequence f by a constant k and, at the same time, scale h by $1/k$ without affecting the noise factor or the correlation product. It is clear, therefore, that we cannot simply use h as a measure of the NEB; our problem is to ‘normalise’ it against a suitable reference. The product

$$\frac{1}{N} \sum_{n=0}^{N-1} |f(nT)|^2 \cdot \frac{1}{N} \sum_{n=0}^{N-1} |h(nT)|^2 \quad (8-25)$$

serves the purpose of eliminating the aforementioned k and, without recourse to a full analysis of the problem, we will adopt this as a likely measure of the ‘normalised noise equivalent bandwidth’ (NNEB) of the CLL.

We can test the validity of this assumption by choosing an M -sequence for f and a conventional ‘early-late’ S-curve response for r ; that is, all values of r are zero except $r(1) = -1$ and $r(-1) = 1$. The 255-bit M -sequence with coefficients $+127/256$ and $-128/256$ has a zero mean, and cross-correlates against a sequence with zero mean and coefficients of zero and ± 4 to give the specified early/late control signal. The three-valued inverse is expected, as it is the sum of two time-shifted bipolar M -sequences. The NNEB is calculated to be almost 2 (actually 508/255). We know that the NNEB term should be unity for a matched system, and we can surmise that the figure of 2 arises because we are essentially performing two correlations (the early and late versions) and so admitting twice the noise power.

For the bipolar $M+1$ sequence discussed in §8.3.3.2 the NEB is 2.726 for the early/late type of control signal. We can attribute this to the fact that the sequence is not ideal. The ‘improved’ code detector results in an NNEB of 85.62. Clearly this requires further investigation. We could surmise that the high value is due to the fact that r being non-zero over most of its range dictates a large value for h . If further analysis confirms this then it would pay to modify the decorrelation sequence h once lock has been achieved.

8.4 Orthogonal Sequences

Although not strictly relevant to the problem under consideration, it is interesting to look at this property, and to ask the question – relevant to communications systems design – of whether, given a set of orthogonal sequences, their inverses are also orthogonal.

Given two periodic sequences f_1 and f_2 we can define them as orthogonal if

$$\sum_{n=0}^{N-1} f_1(nT) f_2(nT) = 0 \quad (8-26)$$

8.4.1 Cases of Correlation

8.4.1.1 Sequences Completely Uncorrelated

If (8-26) were true for all correlation shifts, we would have

$$f_1(nT) \oplus f_2(nT) = 0 \quad (8-27)$$

which is a more severe test than orthogonality. Applying the convolution theorem we can note that for each value of n we have $F_{1,n} \cdot F_{2,n} = 0$. Discounting the trivial situation that one of the functions is identically zero, this means that both F_1 and F_2 must contain zeroes for some elements, therefore it is not possible to calculate their inverses.

We can therefore state that if two functions are completely uncorrelated neither of them has an inverse. This is interesting from a philosophical point of view – two functions have ‘nothing in common’ (because they are completely uncorrelated) and yet, paradoxically, they do share a property after all, because neither of them has an inverse.

One method of generating uncorrelated sequences is to take any sequence, f and to form the double length sequences $[f, f]$ and $[f, -f]$. The former contains only even harmonics in the frequency domain, and the latter only odd harmonics, hence they are uncorrelated in the time domain.

8.4.1.2 Correlation is time-shifted delta function

Looking at a more general case, we have

$$f_1(nT) \oplus f_2(nT) = r(nT) \quad (8-28)$$

where one of the elements of r is zero because the functions are orthogonal. If g_i is the inverse of f_i , then we have

$$\begin{aligned} f_1 \oplus f_2 = r &\Rightarrow F_1 \cdot \tilde{F}_2 = N \cdot R, \quad \tilde{F}_1 \cdot F_2 = N \cdot \tilde{R} \\ f_1 \oplus g_1 = \delta &\Rightarrow F_1 \cdot \tilde{G}_1 = N, \quad \tilde{F}_1 \cdot G_1 = N \\ f_2 \oplus g_2 = \delta &\Rightarrow F_2 \cdot \tilde{G}_2 = N, \quad \tilde{F}_2 \cdot G_2 = N \end{aligned} \quad (8-29)$$

so we can deduce that

$$\begin{aligned}\tilde{F}_1 \cdot \tilde{F}_2 \cdot \tilde{G}_1 \cdot \tilde{G}_2 &= N^2 \\ \Rightarrow G_1 \cdot \tilde{G}_2 &= \frac{N}{\tilde{R}} \\ \Rightarrow g_1 \oplus g_2 &= IDFT\left(\frac{1}{R^*}\right)\end{aligned}\tag{8-30}$$

Now, losing some generality, but making the problem slightly simpler, let us assume that all the elements of r are zero except $r(0) = 1$. That is, $r = \delta$. And we know that $DFT(\delta) = 1$, so this means that $g_1 \oplus g_2 = \delta$ and thus the fact that the f 's form an orthonormal set implies that their inverse, the g 's do too.

Fairly obviously, the δ does not have to be in the zeroth position. For example, if the k th element of r is unity, we have $r = \delta((k-n)T)$, with

$$r = \delta((k-n)T) \Rightarrow R(m\Omega) = e^{-jmk\Omega T} \tag{8-31}$$

and, by implication, we also have

$$r = \delta((k+n)T) \Rightarrow R(m\Omega) = e^{jm(k+n)\Omega T} \tag{8-32}$$

so

$$IDFT\left(\frac{1}{R}\right) = \delta((k+n)T) \tag{8-33}$$

and

$$IDFT\left(\frac{1}{R^*}\right) = \delta((k-n)T) \tag{8-34}$$

because complex conjugation of the DFT is synonymous with time reversal for a set of real coefficients. In summary, we have demonstrated that if the orthonormal set of f obeys $f_i \oplus f_j = \delta((k-n)T)$ for some k then their inverses do too. A quick example using Matlab serves to demonstrate that $r = [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \Rightarrow IDFT(1/R^*) = [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$ as expected.

8.4.1.3 Generalisation

We cannot generalise from the previous example to any value of r because the $1/R$ term means that we cannot sum the coefficients linearly. A simple example serves to show this.

$$r = [0 \ 1 \ 2 \ 3 \ 1 \ 1 \ 1 \ 1] \Rightarrow IDFT(1/R^*) = [11 \ 1 \ -19 \ 51 \ -29 \ 1 \ 21 \ -29]/80$$

without a zero in any position of the sequence.

8.4.2 Summary of Orthogonality Results

- i) if two sequences are completely uncorrelated then neither has an inverse.
- ii) If two sequences are orthogonal and their correlation product is a time-shifted delta function (i.e. similar to the correlation product of a sequence and its inverse), then their inverses will be orthogonal. (But this is something of a special case)
- iii) If two sequences are orthogonal in a general sense (i.e. their correlation product contains a zero at just one shift) then their inverses – if they exist at all – may not be orthogonal except possibly in certain special cases.

8.5 Concluding Remarks

We have presented a simple algorithm for determining the multi-valued inverse of a binary sequence. Under certain circumstances, the inverse does not exist but we explained how this difficulty could be eliminated. We explained the analogy to a matched filter and thereby developed an expression for the signal/noise ratio.

Our intended use for the inverse sequence is to construct a code-locked loop that will allow us to achieve system synchronisation, and to obtain the system impulse response. We observed that it would be possible to design a CLL with no dead-band but, although this could be useful as a method of decreasing the acquisition time, we observed that it degraded the SNR considerably; although we could design a response that had a lower SNR.

Although we intend to make use of this result in the design of the channel sounder, we have not done so yet. The only working part of the equipment is the receiver, with which we have made preliminary measurements of background noise and interference. We will present these in the next chapter, after we have outlined the practical design of the prototype sounder

8.6 References

Al-Dabbagh, A., M. Darnell, A. Noble and S. Farquhar (1997). Accurate System Identification Using Inputs with Imperfect Autocorrelation Properties. *IEE Electronics Letters* 33(17), 1450-51.

Al-Dabbagh, A., T. O'Farrell and M. Darnell (1998). Optical Signal Design Using Reciprocal Periodic Sequences. *IEE Electronics Letters* 34(20), 1962-64.

Blanchard, A. (1976). *Phase-Locked Loops: Application to Coherent Receiver Design*. New York: John Wiley & Sons.

Fan, P. and M. Darnell (1996). *Sequence Design for Communications Applications*. Taunton, England: Research Studies Press.

- Godfrey, K., Ed. (1993). *Perturbation Signals for System Identification*. Hemel Hempstead, UK: Prentice Hall.
- Schwartz, M. (1980). *Information Transmission, Modulation and Noise*. Tokyo: McGraw-Hill Kogakusha. (3rd edition).
- Spilker, J. J. (1963). Delay-Lock Tracking of Binary Signals. In *Phase-Locked Loops and Their Application (IEEE Press Selected Reprint Series, 1978)*, ed. W. C. Lindsay and M. K. Simon. New York: John Wiley & Sons.
- Wolaver, D. H. (1991). *Phase-Locked Loop Circuit Design*. Englewood Cliffs, New Jersey: Prentice Hall.

9 Channel Sounder: Design & Results

Abstract: The preliminary design of the channel sounder is discussed, although the project is still at an early stage and only the receiver has been constructed. The antenna design has not been finalised, but the receiver is able to perform data logging and to take readings of background noise and interference. Results are presented that confirm the operation of the signal-averaging. A complete spectrum from 0-250kHz is presented and discussed.

In the preceding two chapters, we discussed the principles of channel sounding and the sequence design. We explained that it was decided to use a wideband system with a near-PRB sounding sequence and signal-averaging at the receiver. The salient features were

- i) 512 bit sequence generator (modified PRB sequence) clocked at 500kHz (approx. 1kHz spectral lines).
- ii) Data capture at 500kHz (system bandwidth 250kHz): 131072 16-bit samples in 262ms
- iii) 131072-point FFT to obtain spectrum information at 3.8Hz intervals;
7.6Hz noise-equivalent bandwidth.
- iv) Signal averaging of 128 sample sets in 34s.

We shall now discuss the design of equipment to meet these aims, and the preliminary results we have obtained using the equipment.

9.1 Design Evolution

9.1.1 Data Capture

The specification calls for the collection of 16M samples (32Mbytes) per data set, at a rate of $2\mu\text{s}$ per sample. This large quantity of data is perhaps most easily handled by a desktop computer fitted with a high-speed data-capture card. This was considered but, at the time this project was started, it was felt that the transfer of this quantity of data to a hard disc in real time would be a problem, although high data rates such as this are now commonplace. In addition, a PC does not lend itself to portable field-use – even laptop and industrial ‘card’ computers have a very high current consumption and a limited battery life.

It was decided that for the equipment to be suitably portable, a specialised data-capture and logging system would have to be designed. Two approaches were possible – DSP-based or hardware-based. Again, at the time this project was started, it was felt that a hardware-based approach would be the most appropriate, as DSP chips tended to have limited memory and a high current consumption. This decision proved to be a bad one, because of the length of time it took to design and test the hardware. With hindsight, we should have pursued a DSP-chip solution.

The real-time capture of 32Mbyte of data and its transfer to a storage medium was also seen as a problem, although with hindsight this should not have been so. However, it was decided to perform the signal averaging in real-time, in hardware, reducing the dataset to 384Kb so that a single 4Mbit static RAM would be adequate to store the data during this process.

9.1.2 Data Storage & Transmission

Originally it was envisaged that the data-capture unit would contain no storage facility and would transmit the received data along a co-axial or fibre-optic cable to a base station. Cave systems were identified that were suitably extensive, but which had access to mains power close to the entrance, where a desktop computer could be operated. The White Scar Cave system is a show cave in the Yorkshire Dales where such a scheme would be possible. We also identified a number of deep caves in the Yorkshire Dales where field tests could be carried out, although the deepest caves in the UK are in the Brecon Beacon area of South Wales. Further information on suitable test sites is given in appendix A6.

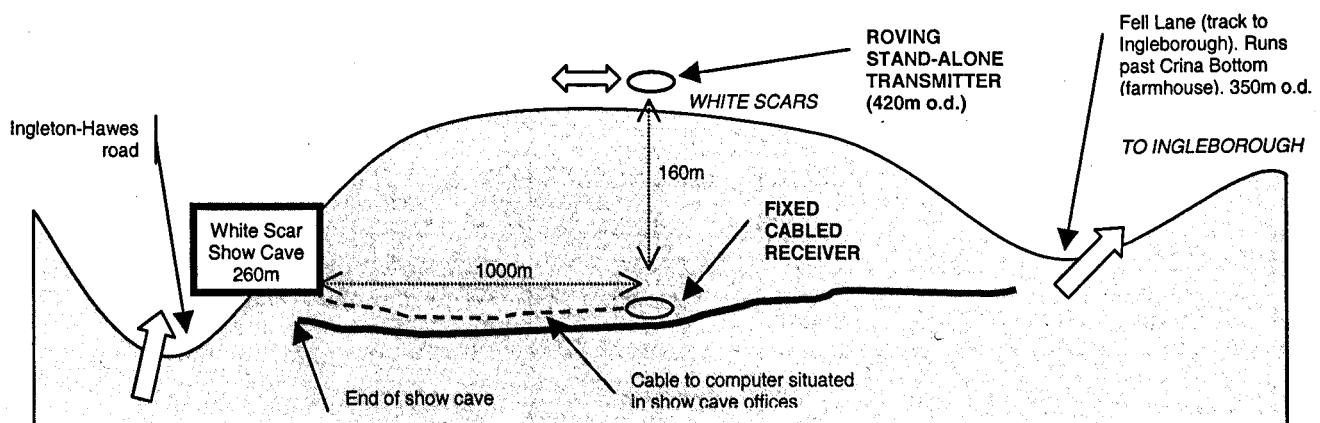


Figure 9-1 : NW-SE elevation of White Scar Cave at grid reference SD 713746.
Possible transmitter and receiver locations for tests requiring mains-powered equipment.

For initial tests at White Scar it was envisaged that the underground receiver would be placed at a far point in the cave system, about 1000m from base, and used with a roving surface transmitter. However, it was felt that the presence of a long copper cable at the receiver could

lead to distortion of the field due to circulating currents, although no evidence has been collected to test that assertion. A simple solution would be to opto-isolate the signal, but although this would have the effect of eliminating earth-return currents, it would not prevent leakage due to capacitive coupling of the cable to the ground. Replacing some or all of the cable with an optical fibre could eliminate the effect of this capacitive coupling. In addition, a high-speed data link would allow the option of transmitting 'raw' analogue data, or digitised data at 8Mbit/s, as well as processed data at a much lower rate. It soon became apparent, when the various schemes were set out on paper, that none were entirely satisfactory, and some were expensive. At the time this project commenced, the best solution seemed to be to transmit the signal-averaged dataset at a low speed using a screened twisted pair, with a section of polymer fibre in the vicinity of the receiver. This scheme would require the receiver to have a certain amount of processing power, which would ideally be provided by an industrial controller.

9.1.3 Microcontroller

As argued above, a microcontroller is needed to control the data capture card, and to provide processing power for the data storage routines. Several devices are marketed, the author being familiar with a product from Triangle Digital Services¹ (TDS) known as the TDS 2020F. Since this unit can interface to a hard disc drive on a PCMCIA card, the problem of local data storage is solved and onward transmission of the data is no longer essential. This leads us to the possibility of using the sounder for long-term data logging, which will be the next phase of the development.

The TDS 2020F is an industrial programmable controller that executes the language ANS-Forth on a Hitachi H8 16 bit CPU. Forth is a fourth-generation object-oriented language that is ideal for this task, although it is an extremely specialised language. The programming environment is unlike that of more conventional PC systems, because compilation takes place on the 2020F board, with the PC acting solely as a terminal and file storage system. The author has considerable experience of Forth, including the writing of Forth compilers and interpreters and the use of Forth in industrial control software.

An extensive range of library modules are available from TDS, which allow high-level interface to a number of peripherals including, for example, GPS cards and modems. A number of peripheral modules are available for the TDS 2020F, including a PCMCIA interface that will accept a 500Mbyte miniature hard disk. Being designed for portable datalogging, the current consumption is very low, with the manufacturers claiming superiority over conventional industrial PC cards.

¹ www.TriangleDigital.com

We now have the situation that the data is averaged in real-time in hardware, for speed, and stored on a local hard disc with all operations being under the control of a specially programmed microcontroller. The receiver is fully portable, but can be left on site for long periods, with the data being stored locally for later retrieval, or transmitted at a low rate on opto-isolated copper twisted pair or cheap polymer fibre.

9.1.4 Receiver synchronisation

The transmitter and receiver clocks need to be synchronised if there is to be any benefit from the signal-averaging process. We described, in the preceding two chapters, how this can be achieved using a code-locked loop.

9.2 Final Design

The channel sounder has evolved into a stand-alone recorder that can be programmed using a keypad and LCD to make individual readings on demand, or to make a series of readings, in the fashion of a data-logger, over a period of many months. Synchronisation will be by means of a code-locked loop. We also envisage that each unit, containing a transmitter and receiver, will be able to transmit and receive data, and so a transmitter/receiver pair could undertake a dialogue as part of the sounding process. However, we have also realised that the design of the sounder is a larger project than was first envisaged and we have not progressed beyond preliminary trials of the receiver section.

9.2.1 Status of Design

Circuit diagrams and further design details are given in appendix A7.

9.2.1.1 *Receiver*

A prototype receiver has been constructed, comprising a motherboard for the TDS 2020F controller, and a DSP processing board that contains a receiver pre-amp, the analogue to digital converter (ADC) and the hardware signal-averaging circuitry. The 2020F controls the operation of the DSP card, and the transfer of the data from its temporary storage in RAM to the local PCMCIA-based hard disc drive. There is also a data display (LCD) that reports status information to the user.

The design is not complete and, in particular, we have not written the software that would allow the unit to operate in a stand-alone fashion with its own keypad or keyboard. Instead, it is controlled through a serial link from a laptop PC. The other two major areas where the design is not complete are

- i) Receiver pre-amp and automatic gain control.

We intend to design a high-gain programmable amplifier that is correctly noise-matched

to the antenna. At the moment, a low-gain pre-amp on the DSP card is being used for tests.

ii) Code-locked loop.

We have not designed the code-locked loop circuitry. This will be a plug-in board that will over-ride the existing fixed-frequency 16MHz master clock. In part, the CLL feature depends on the transmitter circuitry being present (because the same circuit modules are used in both instances) and this has not yet been designed.

In conclusion: the receiver section of the channel sounder is operational only in a 'laboratory' setting (i.e. with a laptop PC acting as a terminal) and with a low-gain non-matched pre-amp. In this state the device cannot be used for channel sounding although it can, with suitable programming, be used as a stand-alone logger of background noise and interference. Some preliminary results are given in later in this chapter.

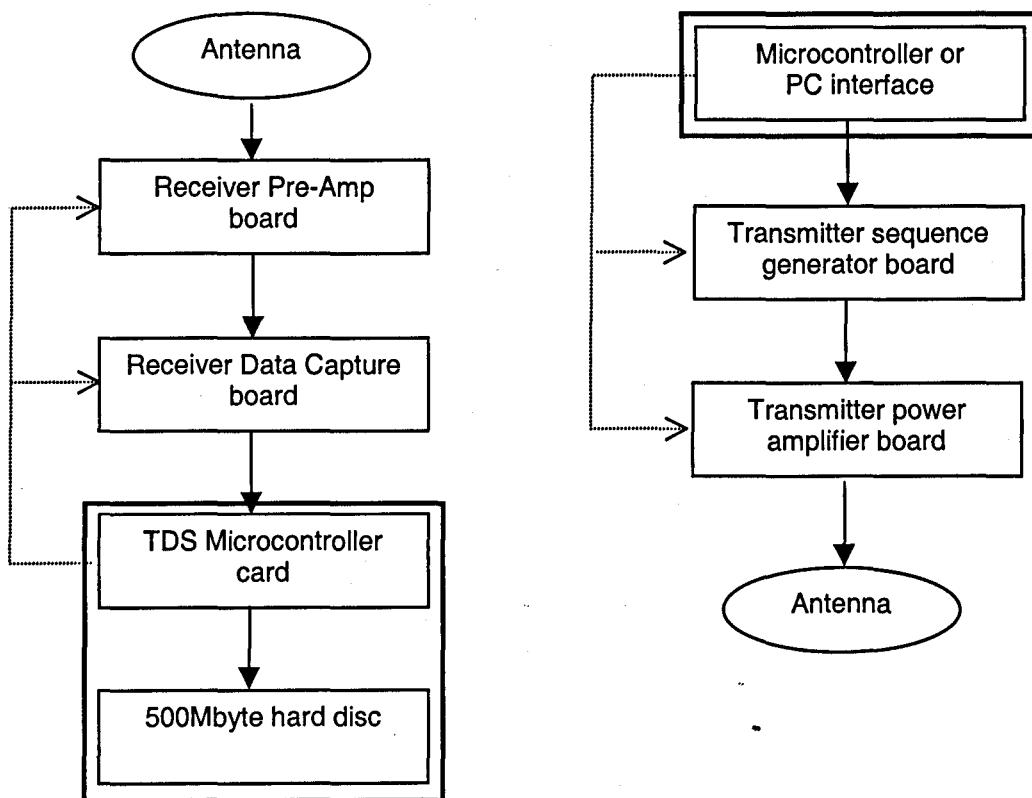


Figure 9-2 : Block diagram of prototype channel sounder.

9.2.1.2 Transmitter

We designed and built a transmitter based around one of the PIC range of microcontrollers from Microchip². The microcontroller was programmed with a look-up table containing the modified PRB sequence that we intended to use. The microcontroller outputs controlled a half-bridge circuit similar to that shown in *Figure 5-3*. The bridge operated from a single +12V power rail, and the antenna load floated with the aid of a 10µF 25V low-loss multi-layer ceramic capacitor of the type used in switched-mode power supplies. This prototype transmitter was built to test some of the concepts and was not intended to be used for measurements; the main problem being that the MCU would not operate fast enough to clock the data at 500kHz.

The channel sounder described above is a modular device, and we considered that it would be advantageous for the transmitter section to be constructed as an additional module, that could utilise the same microcontroller routines. However, although preliminary design sketches have been made of the transmitter, no further progress has been made with the design.

9.3 Evaluation of Prototype

9.3.1 Programming the TDS controller

The procedure for compiling and running Forth programs is somewhat different to the method used with other high-level language-controllers. In this instance, the Forth operating system runs interactively on the TDS board, with a PC acting solely as a terminal and file storage system. Programs are stored as text files on the PC and uploaded as if they were being typed interactively. All compilation takes place on the TDS board and the resultant program module can be ‘burned’ into flash memory for stand-alone execution.

This procedure means that, for development work, the user can type commands and execute Forth routines interactively from the PC terminal, without having to compile a stand-alone ‘turn-key’ program after each modification. The disadvantage is that the MCU requires a PC terminal for it to operate. Because the software is still under development, we did not attempt to write a turn-key system and, instead, controlled the TDS 2020F and the channel sounder by interactive commands typed at the PC keyboard. The simple routine we used to collect the background noise data also executed the following sequence of actions.

- i) Reset the channel sounder; instruct the ADC to self-calibrate
- ii) Program the number of datasets for signal-averaging

² www.microchip.com

- iii) Wait 10 seconds to give the operator time to power-down the PC (and so remove a severe source of interference)
- iv) Collect and signal-average the requested number of datasets
- v) Write the data to the hard disc
- vi) Announce that the process has finished.

After all the data was collected, the hard disc was removed and plugged into the PCMCIA slot of the PC. The data was then copied to the PC and processed using Matlab, producing the graphs shown later in this chapter.

In a field environment, the TDS card could be programmed to collect data over a period of several months, and the hard disc card could be retrieved and replaced at suitable intervals.

9.3.2 Data Processing

9.3.2.1 A to D Converter

As we have described previously, the receiver samples the signal at 500kHz, capturing 2^{17} samples per dataset and resulting in spectral data at 3.8Hz intervals. The 16-bit ADC was an AD7721 from Analog Devices, featuring an over-sampling anti-aliasing digital filter with a 3dB bandwidth of 244kHz (16MHz clock) and an attenuation of 72dB at 276kHz. Each reading of the complete spectrum from 0–250kHz took 262mS.

The first image frequency of the ADC's over-sampler is quoted as being 0.986 times the clock frequency, i.e. about 15.8MHz. The antenna pre-amp had a first-order RC low-pass filter at 201kHz, resulting in an attenuation at the image frequency of $1/\sqrt{1 + (f_i/f_c)^2} = 38\text{dB}$.

There was also a first-order filter at 185.5kHz in the pre-amp; and a high pass filter to remove some mains interference below 190Hz. For these preliminary results, the receiver antenna was not calibrated but the response was believed to be substantially flat response between the filter limits set out above.

9.3.2.2 Fourier Transform & Windowing

The data was processed using a 131072-point discrete Fourier transform (DFT) using the MatLab software package. When calculating the Fourier transform of a finite sequence it is implicit that the sequence can be extended to infinity in a periodic fashion. But this process can lead to discontinuities at the bounds of the sequence, which have the effect of introducing additional spectral components into the signal. This phenomenon is often referred to as *spectral leakage*. It can be reduced by decreasing the sampling period, although the leakage is not caused by the sampling process itself.

The solution to spectral leakage is to apply a *window function* to the data, to smooth the signal transitions at the edges of the sequence. Use of a windowing function increases the

resolution of the frequency spectrum, but it does this at the expense of broadening the spectral lines slightly. Normally this is not a problem, but we wanted to investigate the data for evidence of synchronised signals, which would have a spectral width of just one point, so it was also necessary to study the un-windowed data.

A 131072-point waveform, sampled at 500kHz has spectral lines at 3.815Hz intervals. Without windowing, the resolution bandwidth (the 6dB bandwidth³) of each point in the frequency spectrum is 1.21 ‘points’, i.e. 4.6Hz. However, the sidelobes are as high as -13dB so there is considerable ‘clutter’. The noise equivalent bandwidth is one ‘point’ or 3.8Hz.

Windowing was implemented using a modified 4-term Blackman-Harris window, which is described in appendix A8. This has a resolution bandwidth of 2.67 points, or 10.2Hz, but a maximum sidelobe height of -90dB, so this is an ideal window to use. The window has a coherent gain of 0.36 times that of a rectangular window, and this is taken into account in the plots that are presented here. The noise equivalent bandwidth is two ‘points’ or 7.6Hz.

9.3.3 Antenna Design and Calibration

We made no attempt to calibrate the antenna for these preliminary tests, nor to apply, in detail, the design rules discussed in earlier chapters. The antenna was constructed from 50 turns of 7/0.2 PVC-coated copper equipment wire, which was wound onto a 450mm diameter former cut from a 50mm sheet of expanded polystyrene. The calculated specific aperture of such an antenna is $3.4\text{m}^2/\sqrt{\Omega}$ and the bandwidth 1200Hz. The bandwidth is reduced by the large damping resistor but, in the context of an untuned antenna, the bandwidth is simply a confirmation that we are operating in the region where the antenna is acting as an differentiator.

The induced voltage is $\omega\mu H \cdot NA$ so we can see that the theoretical ‘transfer gain’ of the antenna is approximately $63\mu\text{V}/\text{Hz}/\text{Am}^{-1}$, or $50\text{nV}/\text{Hz}/\text{nT}$. The pre-amp gain is 40 so, for example, at 50kHz the response of the antenna is $130\mu\text{V}/\mu\text{Am}^{-1}$ or $100\text{mV}/\text{nT}$. A field strength of $8\mu\text{A}/\text{m}$ (a flux density of 10pT) would result in an ADC reading of 1mV.

9.4 Analysis of Results

9.4.1 Self-Noise

Operating the prototype channel sounder receiver with a shorted input resulted in an r.m.s. signal voltage of 1.77mV, (*Figure 9-3a*) corresponding to a mean spectral density of $3.55\mu\text{V}/\sqrt{\text{Hz}}$, which was higher than expected. The rise in signal level is thought to be an equipment problem, and will be investigated. There was a significant noise contribution at

³ Not, as we would usually expect, the 3dB bandwidth. This is explained in appendix A8.

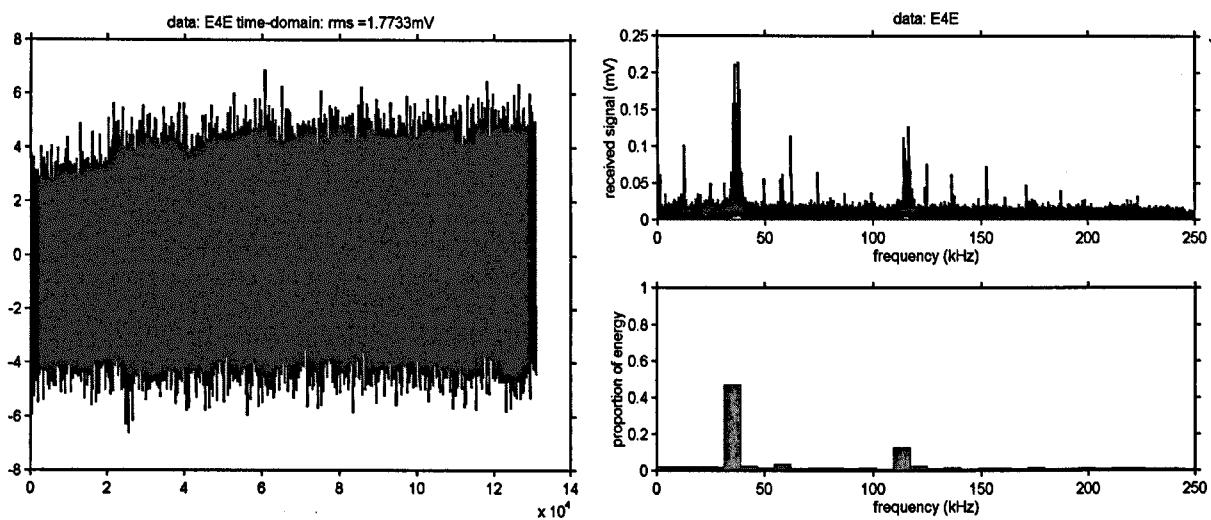
35–39kHz and another at 113–118kHz (*Figure 9-3b*). This broadband noise could arise from the switch-mode power supply in the LCD module that was present on the TDS 2020F motherboard.

9.4.2 Signal Averaging

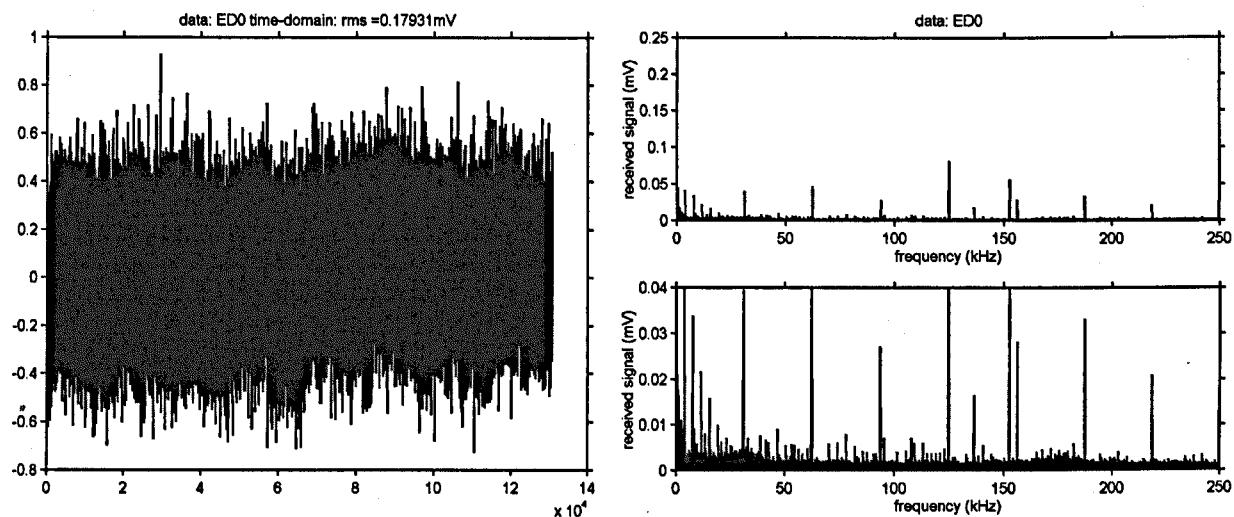
Applying signal-averaging over 100 datasets reduced the noise level by a factor of almost ten, as would be expected, to 0.18mV and $0.36\mu\text{V}/\sqrt{\text{Hz}}$. (*Figure 9-4a*). The effect of signal averaging is clearly demonstrated in *Figure 9-4b* where the broadband noise peaks, mentioned above, were reduced significantly. Some noise spikes remained unaltered – for example the spike at 125kHz. It is obvious why this should be the case – an inspection of the un-windowed data showed that they were a single sample wide, and that they were at exact sub-multiples of the clock frequency of 500kHz, thus demonstrating that they arose from internal noise synchronised to the counter chain that controls RAM access.

The presence of these noise spikes will affect the ability of the instrument to perform channel sounding, so this will have to be investigated. But it was not surprising that there was some interference because neither the test antenna nor the data capture module was at all shielded. However, the persistence of the spikes after signal averaging clearly demonstrates the principle of the channel sounding – namely that the synchronised signal is enhanced at the expense of channel noise.

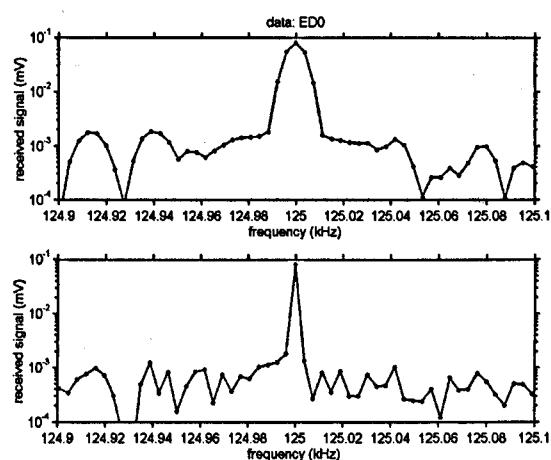
Investigation of the interference spikes requires un-windowed data, because the presence of the window would tend to broaden the noise spikes slightly. With no window applied, a spike that is a single sample wide can almost certainly be assumed to be synchronised to the sample frequency. With a window applied, it might not be possible to tell the difference between a synchronised signal and an interfering signal that was very close in frequency. This is significant because TV timebase interference is nominally at a sub-multiple of our clock frequency. As an example, consider the 125kHz noise spike shown in *Figure 9-5*. The un-windowed data (bottom) shows that the spike is a single ‘point’ wide. (The data points on either side are more than 30dB down). The top graph shows the windowed data, where the peak is substantially broadened. Notice that the windowing function allows the signal to be extracted from the general noise.



*Figure 9-3 : Self-noise recorded with shorted input.
a) Left: Y-axis: signal (mV). X-axis: sample. b) Right: Freq. spectrum (windowed) and energy histogram.*



*Figure 9-4 : Self-noise with signal averaging of 100 datasets.
a) Left: Time sequence. Y-axis: signal (mV). X-axis: sample. b) Right: Frequency spectrum (windowed).
Top Right: the scale matches Figure 9-3. Bottom Right: enlarged scale.*



*Figure 9-5 : Enlargement of Frequency spectrum.
An enlargement of Figure 9-4, (windowed (top) and un-windowed (bottom)).*

9.4.3 Spectrum Survey: 3 orthogonal axes

The receiver section of the channel sounder was used to undertake a preliminary spectrum survey. Three orthogonal readings were taken with the test antenna, for vertical field, north-south field and east-west field. The N-S and E-W fields were so named for convenience – in practice, the antenna was aimed at 197° magnetic (roughly SSW) and 107° (ESE). 197° is the direction of the BBC Radio 4 transmitter (198kHz, 500kW) at Droitwich. This is some 180km from Leeds⁴. The vertical and E-W measurements were made on Sunday 9th June at 13:30 BST. The NS measurement was made on Monday 10th June.

The Sunday measurements showed a strong signal at 15.625kHz and its harmonics. These are sub-multiples of 500kHz but, unlike the self-noise signal mentioned above, the spectrum of these lines is slightly smeared thus demonstrating that they must originate (for the most part anyway) from an unsynchronised signal; this being the TV timebase for the UK's 625-line PAL television system. In fact, the self-noise test showed no significant contribution at 15.625kHz at all.

9.4.3.1 Vertical Field

The received signal was around 15.7mV and predominantly 50Hz (*Figure 9-6a* on page 190). The spectrum is dominated by the timebase interference at $F(4096)^5$, which is 8.2mV at 15.625kHz and by mains interference, with sample $F(13)$, corresponding to 49.6Hz, with a level of 15.6mV. (*Figure 9-6b*). This demonstrates that some more severe mains filtering may need to be included in the pre-amp to improve the dynamic range, although the mains interference in the field is likely to be much lower than in these, preliminary, lab. readings. An analysis of the frequency spectrum showed that some 63% of the signal energy lay in the lowest 7.8125kHz band, and some 10% in the third such band (which includes the 15.625kHz line at its lowest end).

9.4.3.2 Analysing the timebase peak

An inspection of the 15.625kHz spectral line serves to demonstrate some of the analysis we can perform. First of all, (*Figure 9-7a* on page 211) we note that the un-windowed data shows the data points on either side of the spike to be as high as -12dB, so the spike probably originates from an un-synchronised signal. Next, we note that the windowing function broadens

- 4 Leeds is grid ref SE3034, Droitwich is SO8962. In absolute grid co-ordinates, in km, Leeds is therefore 130234, Droitwich 089062. The vector distance is therefore (041, 172) giving a direct distance of 177km and a bearing of 193° (grid). Mag. N is currently 4°W of true north, and grid north more-or-less coincides with true north at Leeds, so Droitwich is. 197° mag.
- 5 $F(4096)$ denotes the 4096th sample in the sequence F , which is that of the frequency spectrum. Counting begins at $F(0)$ which is the d.c. level. There is potential for confusion here, because Matlab arrays begin counting at 1, not zero, so $F(4096)$ corresponds to index 4097 in the arrays in Matlab.

the peak and also has the effect of resolving the ‘clutter’ underneath the peak, to what looks like sidebands.

At a higher magnification (*Figure 9-7b*) this is seen to be the case, and so we can deduce that the 15.625kHz signal was modulated by a 50Hz signal. This lends further evidence (as if any was needed) that this is a TV timebase signal. The spectrum plot is of the magnitude of the DFT. If we were to investigate the phase of the DFT it might allow us to draw conclusions as to whether the modulation was of phase or amplitude although this is not relevant to the task in hand.

Below a few kHz, individual mains harmonics were evident (although not at the scale of this plot). The test antenna did not fully filter the 50Hz mains interference which was one reason why the gain of the antenna had to be kept low. The significant harmonics are mainly the odd ones. The UK cave radiolocation frequency is usually 874Hz. The alternative beacon frequency, at 3496Hz is not troubled by any mains harmonics at the scale of this test. The peaks above 0.4mV in the range 0–1kHz are shown in *Figure 9-8* and are listed in *Table 9-1* below.

50Hz	16.2mV	99Hz	1.8mV
149Hz	3.1mV	202Hz	0.6mV
248Hz	0.9mV	301Hz	0.8mV
351Hz	4.2mV	401Hz	0.8mV
450Hz	1.5mV	493Hz	1.0mV
652Hz	1.5mV	752Hz	0.7mV
851Hz	1.1mV	950Hz	0.9mV
1152Hz	0.7mV		

Table 9-1 : Values of mains harmonics, from Figure 9-8.

9.4.3.3 E-W Field

The EW plot was, likewise, recorded at a time when TV timebase interference was high, but the horizontal field at *F* (4096) is slightly lower, at 4.9mV. The 50Hz line is lower, at 11.2mV but, despite this, the rms signal voltage was 50mV with peaks at almost 150mV (*Figure 9-9* on page 212). The spectrum shows that was largely due to the broadband signal from 109–111kHz that contains 86% of the signal energy.

9.4.3.4 N-S Field

The most interesting plot is the NS result. (*Figure 9-10* and *Figure 9-11*). This spectrum was recorded during a weekday when there was less TV interference. It was noticeable that unless all nearby computer equipment was turned off there was considerable additional interference present during the lab. readings notably timebase signals at harmonics of 48.4kHz from a high-resolution TV monitor.

The test antenna had only a low gain, so the sidebands of some signals were partially buried in noise. Nevertheless, a clear 198kHz carrier could be observed, as well as the 60kHz carrier of the NPL's Standard Time transmission from Rugby and several other identifiable stations.

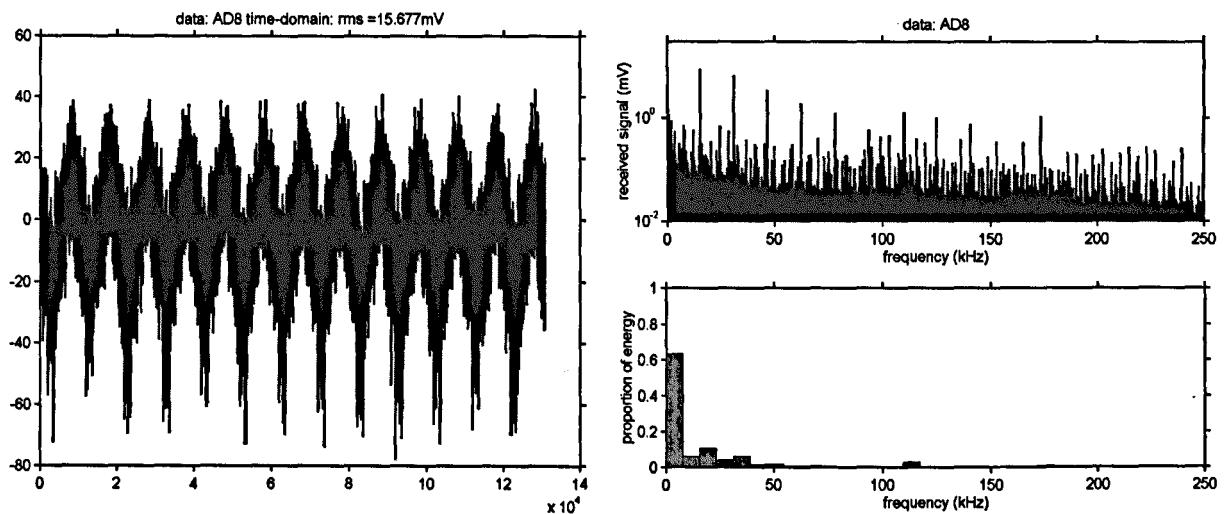
The rms signal voltage was 19mV, with 17% of the energy in the lowest 8kHz (nominal) band, 3% in the broadband signal from 94–102kHz and 72% in the 195–201kHz band encompassing the BBC R4 signal.

The broadband signal at 96–102kHz shows a comb-like structure and is believed to be one of the Loran navigation beacons, although this has yet to be confirmed.

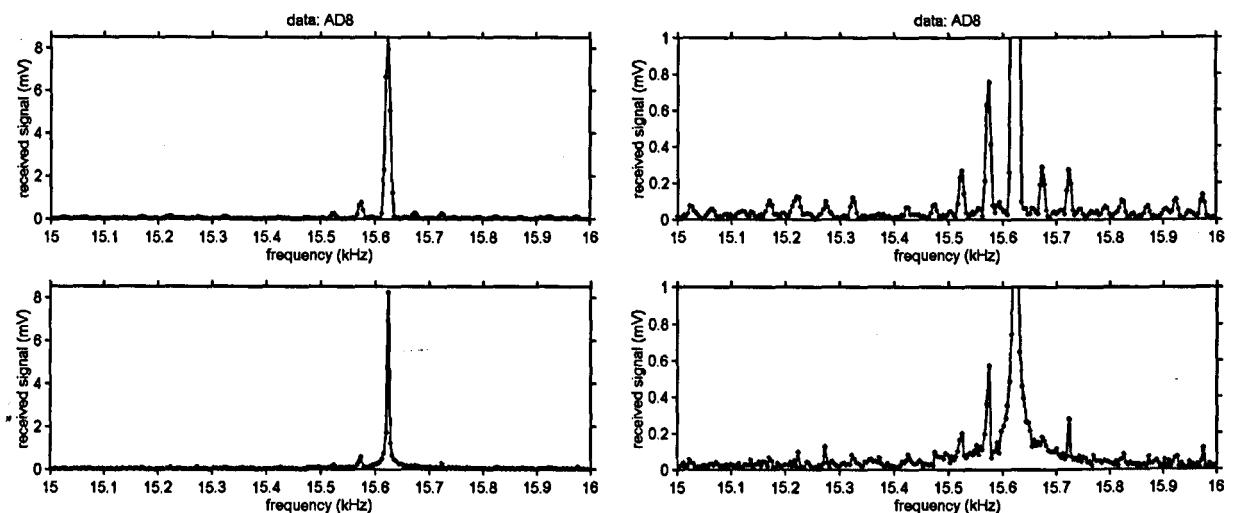
Figure 9-11, on page 212, shows an enlarged section of the frequency spectrum, and demonstrates that the energy in the band is more or less constant, with a few spectral lines superimposed on it. For example, at 85kHz, each of the bars in the figure represents between 10 and 20 p.p.m. of the energy in the spectrum. The r.m.s. voltage (*Figure 9-10a*) is 19mV, so the r.m.s. voltage of one of the histogram bars would be $19\text{mV} \times \sqrt{10 \text{ p.p.m.}} = 60\mu\text{V}$. The bandwidth of each bar is 244Hz, so the noise density is $3.8\mu\text{V}/\sqrt{\text{Hz}}$. Note that this is only slightly higher than the self-noise measured earlier, which suggests that some attention to the antenna and pre-amp screening is required (because the atmosphere, not the receiver, should dominate the noise in a well-designed receiver).

Notice that – in this test at least – the spectrum from 40kHz to 100kHz seems to be fairly quiet apart from the major signal peaks. This may be a false assumption, caused by the pre-amp noise masking the weaker interfering signals. Until the antenna is properly designed and calibrated we should not place too much emphasis on this.

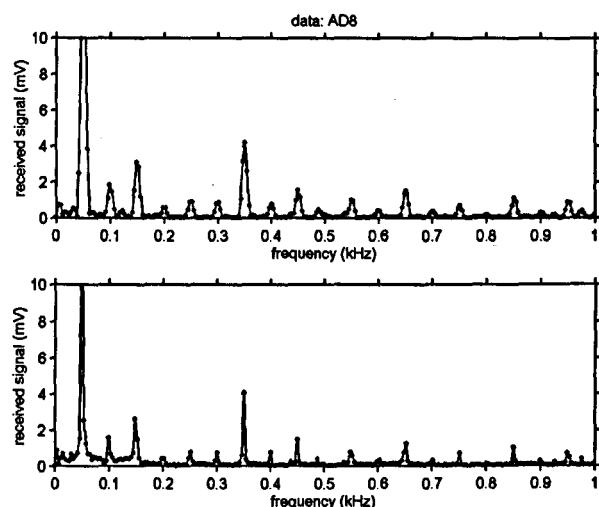
The noise voltage density of $3.8\mu\text{V}/\sqrt{\text{Hz}}$ at 85kHz can be converted to an equivalent magnetic noise field spectral density of $710\text{nA/m}/\sqrt{\text{Hz}}$ (-123dB) by the transfer gain of the antenna, as derived in §9.3.3. We can then use the results of §6.2.1 to derive an atmospheric noise temperature ratio, F_a , of 168dB for the specific case of a north-south-orientated elemental horizontal magnetic dipole. The CCIR/IRU-published figure for F_a at 85kHz is between 80 and 139dB (see *Table 6.1* in §6.2.2) and a typical value for man-made urban noise is 100dB (from ITU report). We discussed, in §6.2.1, that the published figures for F_a could not necessarily be taken as valid for the magnetic field. But here we have an example of a value for F_a that has specifically been derived from a magnetic antenna. A program of future work is planned to collect further data using magnetic antennas.



*Figure 9-6 : Vertical field. Time and frequency spectrum.
a) Left: Y-axis: signal (mV). X-axis: sample. b) Right Top: Spectrum (windowed), Right Bottom: energy histogram.*



*Figure 9-7 : Frequency spectrum, vertical field, showing TV timebase.
Top: windowed data, bottom: un-windowed data.*



*Figure 9-8 : Frequency spectrum, vertical field, showing mains harmonics.
Top: windowed data, bottom: un-windowed data.*

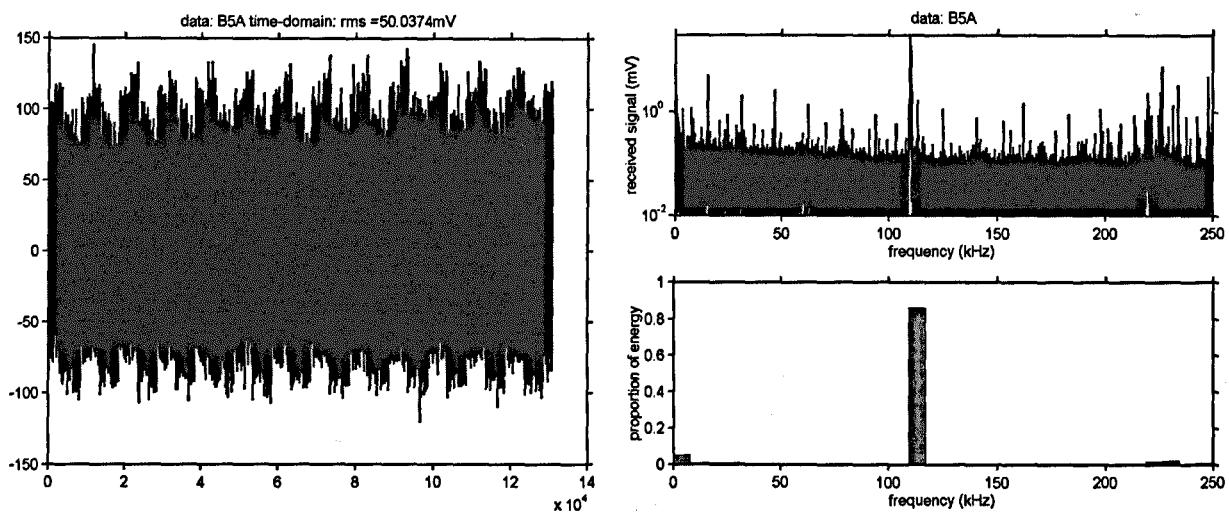


Figure 9-9 : Horizontal E-W field. Time and frequency spectrum.

a) Left: Y-axis: signal (mV). X-axis: sample. b) Right Top: Spectrum (windowed), Right Bottom: energy histogram.

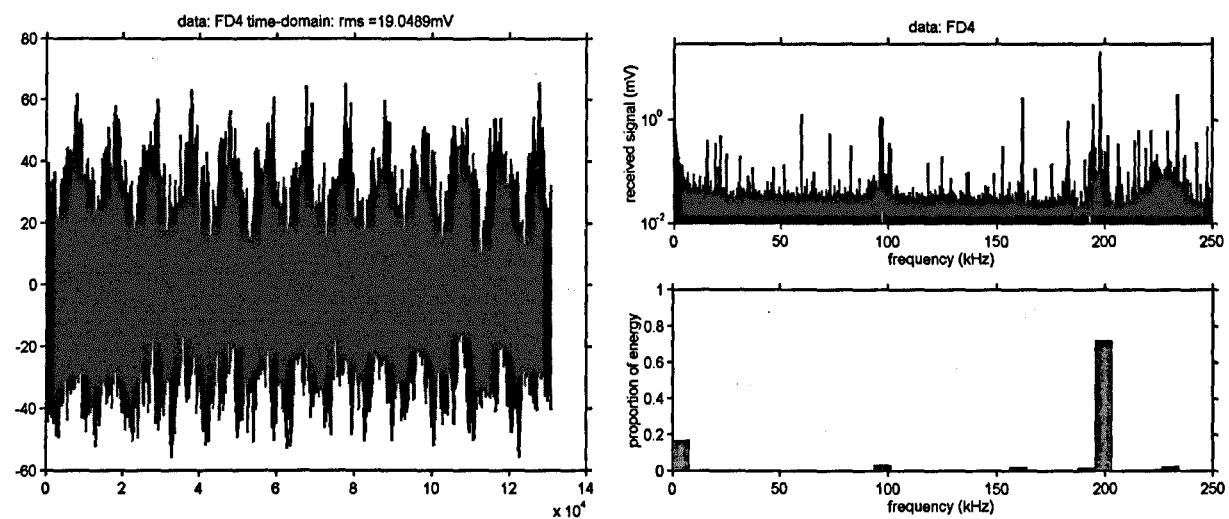


Figure 9-10 : Horizontal N-S field. Time and frequency spectrum.

a) Left: Y-axis: signal (mV). X-axis: sample. b) Right Top: Spectrum (windowed), Right Bottom: energy histogram.

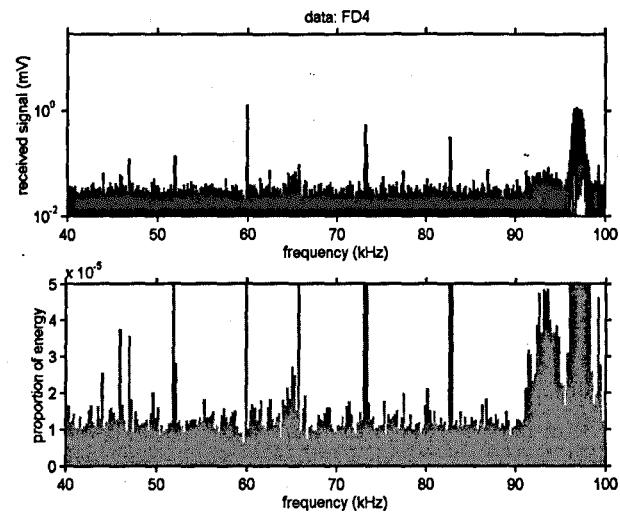


Figure 9-11 : Horizontal N-S field. Frequency spectrum 40-100kHz.

Top: Frequency spectrum (windowed), Bottom: energy histogram with bars 244Hz wide.

9.4.4 A Detailed View of the Spectrum

The N-S data is now given in an enlarged form, with each graph covering 12.5kHz of the spectrum, together with some brief observations. At the scale of these plots, with enhanced line thickness, it is still difficult to see close detail.

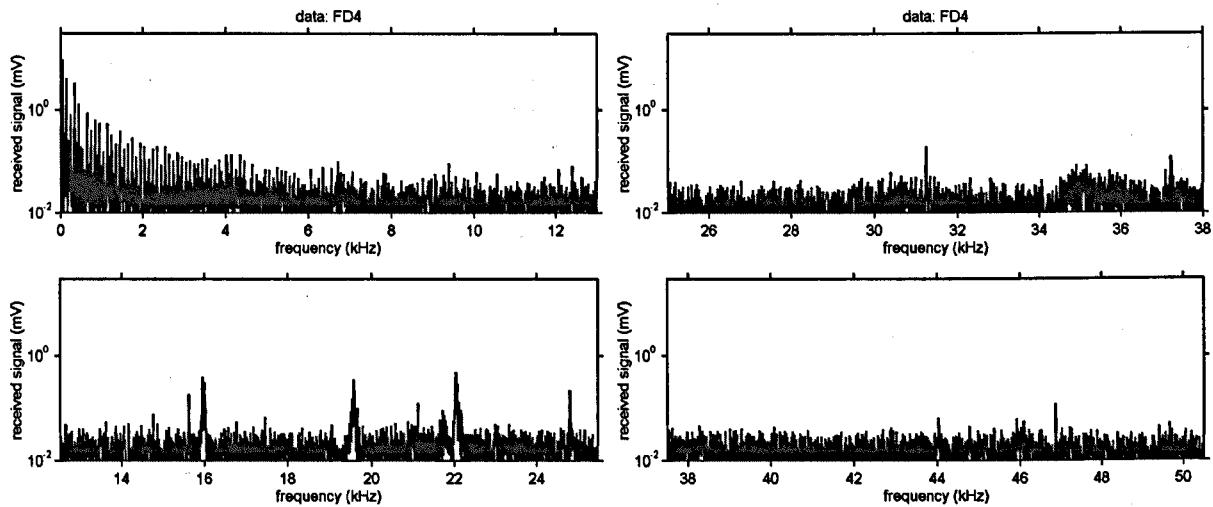


Figure 9-12 : Horizontal N-S field. Frequency spectrum 0-50kHz.

0-50kHz: The TV timebase line at 15.625kHz can be observed, and a slightly wider signal at 16.00kHz, which is probably the 16kHz standard time signal. Peaks at 19.59kHz and 22.06kHz could be a military station, which was set up recently, and which has caused problems for commercial pipe location devices that operated on 22kHz. There are also peaks at 24.83kHz and 31.25kHz, the latter being a TV timebase harmonic.

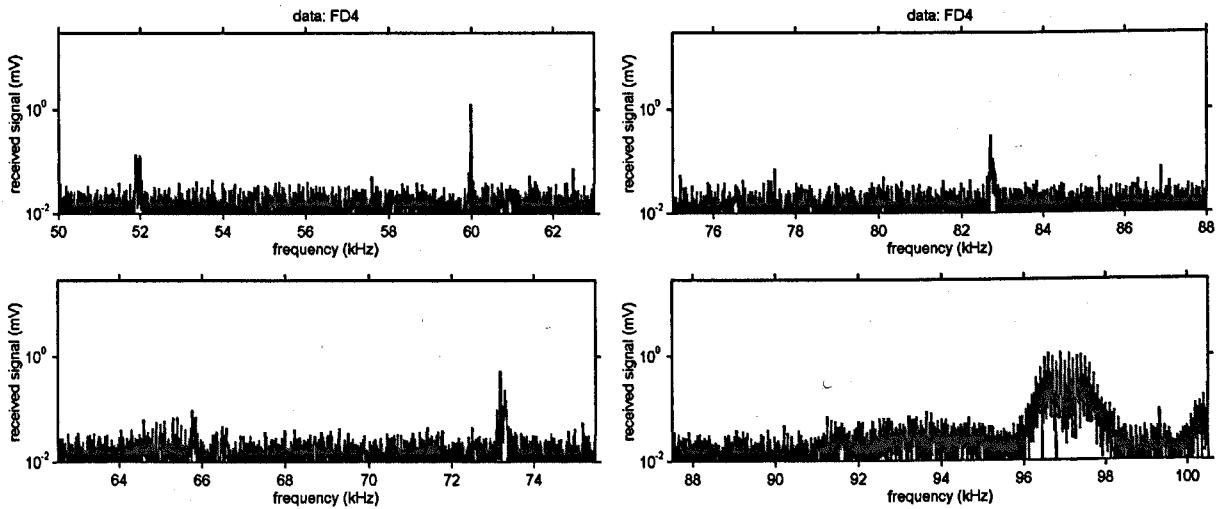


Figure 9-13 : Horizontal N-S field. Frequency spectrum 50-100kHz.

50-100kHz: There are a pair of peaks at 51.90 and 52.0 kHz, suggesting that this could be a 50Hz data transmission using BPSK or DSBSC modulation. The 60kHz standard time signal at

Rugby is visible. There is a 200Hz-wide signal at 73.20kHz which, according to Chris Trayner (*pers. comm.*), "carries encrypted data in a peculiar modulation format" and is possibly located in the Rugby area. Most of the narrowband signals in the LF spectrum are teleprinter and data transmissions. A peak at 82.71kHz is close to the UK cave rescue frequency. The largest signal under 100kHz is the broadband signal from 96.1kHz to 98kHz. A closer view of this shows the comb-like structure we would expect from the Loran beacons.

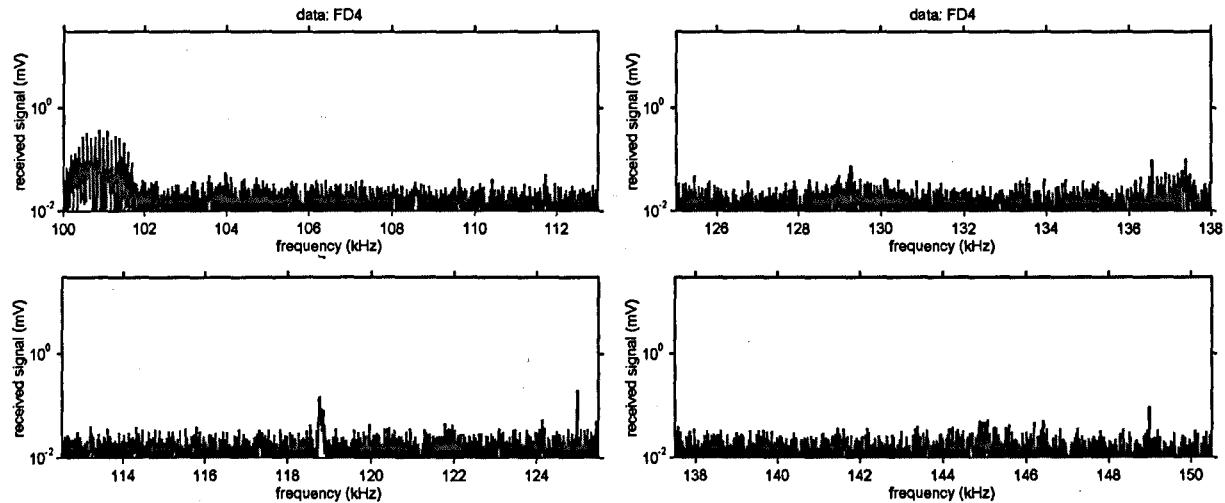


Figure 9-14 : Horizontal N-S field. Frequency spectrum 100-150kHz.

100–150kHz: The broadband signal from 100.2kHz to 101.6kHz is probably also a Loran transmission. We have not yet tried to match any of these signals to the list of Loran beacon frequencies. There are narrow peaks at 118.77kHz, and 125.00kHz. The 125kHz spike is probably internal clock.

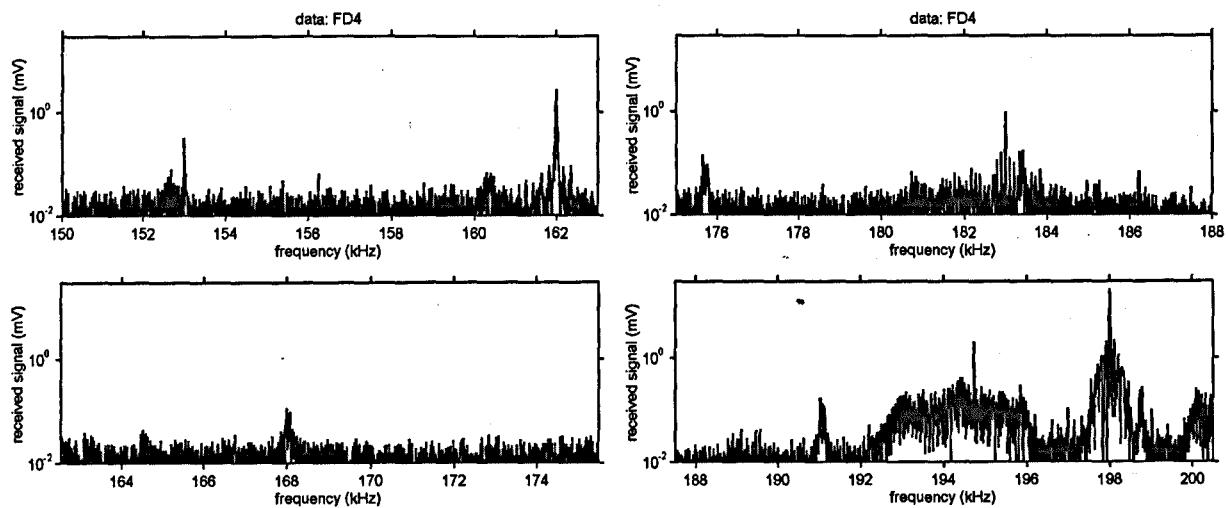


Figure 9-15 : Horizontal N-S field. Frequency spectrum 150-200kHz.

150–200kHz: There is a narrow peak at 153.00kHz. This is a multiple of 9kHz, which is the standard channel spacing in the LF band. At 162.00kHz there is clear evidence of sidebands to the signal. Peaks at 168kHz and 175.66kHz are followed by another clearly visible modulated carrier at 183kHz (which is not a multiple of the 9kHz channel spacing). There is another 30Hz peak at 191kHz and then what can only be the sidebands of the powerful BBC R4 broadcast station with its carrier at 198.00kHz. John Rabson (CREG journal editor) reports (*pers. comm.*) that

"153 kHz is used by Deutschlandfunk. 162 kHz is France Inter. 183 kHz is Europe No 1 (a French language commercial station located in Germany). When the LF and MF broadcast channels for region 1 were adjusted to give carrier frequencies that were a multiple of 9 kHz this station did not follow suit. The same also applies to 177 kHz (Oranienburg in Germany)".

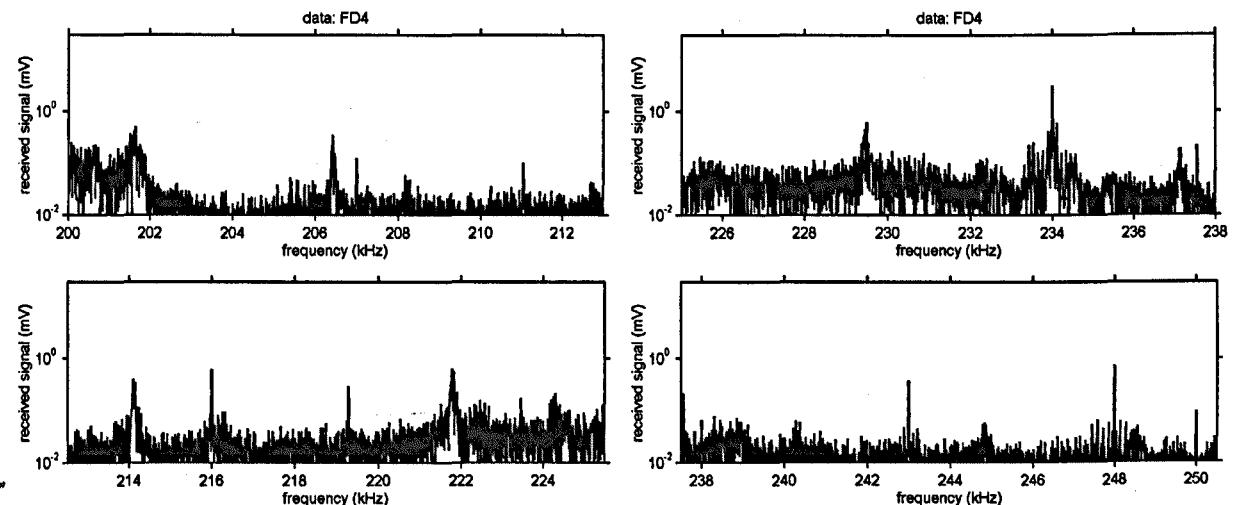


Figure 9-16 : Horizontal N-S field. Frequency spectrum 200-250kHz.

200–250kHz: This section is dominated by modulated carriers on the 9kHz channel spacing – 207kHz, 216kHz, 234kHz, 243kHz. In addition there is what appears to be a modulated carrier at 206.4kHz, 214.14kHz, 219.30kHz, 221.79kHz, 229.3kHz, and 248kHz, but we have not investigated closely.

9.5 Concluding Remarks

We have presented preliminary results from a laboratory test of the receiver section of the sounder, in which it logged the background noise and interference. An accurate calibration was *not* carried out, and so the results are mainly useful as a confirmation that the equipment was functioning. Although it was possible to identify broadcast radio stations, it was not possible, in this brief test, to quantify the noise for different antenna orientations due to the high levels of mains and television interference in the domestic location used for the tests.

The next stage of the development should be to undertake a more comprehensive noise analysis – looking at different antenna orientations. This may have to be undertaken at a quieter rural location. Following this, the equipment must be further developed so that we can undertake a full sounding exercise by transmitting a synchronised wideband sequence.

Eventually, the data-logging equipment will be programmed to take automated measurements at a remote underground location over a period of several months. This work had to be postponed in 2001 because the outbreak of Foot and Mouth Disease in the UK made access to caves impossible for most of the year.

It is known that atmospheric noise varies with the seasons, and we expect that there might also be a variation in underground propagation, due to the changing water content of the peat overburden in many limestone areas of the UK.

10 Concluding Remarks & Recommendations for Further Work

10.1 Propagation

In *chapter 2* we analysed the field expressions using formulations originally derived by Wait in the 1960s. Desktop computers can now facilitate a level of investigation not previously possible and, although we limited our discussion to an analysis of the vertical magnetic dipole (VMD) in a coaxial configuration, we described a total of eight independent antenna configurations that would benefit from further analysis and simulation. The simulations demonstrated some unexpected field behaviour, although we also showed that, for distances close to the field source, an analytical solution based on the assumption of an infinite medium gave acceptable results.

We observed that a detailed study of ground conductivity, with particular reference to sub-surface communications in limestone, did not appear to exist; and we suggested that this would be a suitable research project. The investigation would be of benefit to the study of radiolocation – used for mines and cave rescue, civil engineering and cave surveying – as well as producing results of interest to geophysicists.

We commented only briefly on radiolocation, noting that the results of the computer field simulation had considerable application in this area, where we had already published original material explaining how the inherent inaccuracy in location at skin-depth distances could be overcome. Future studies in this area could attempt to combine our results with the location work by Nessler, and the result could be an enhanced system of personal location for trapped miners.

We described the concept of an optimum frequency, and how this arose as a result of the combination of the principle of magnetic induction and that of skin depth attenuation. The theoretical result was that the optimum frequency for coaxial communications with a VMD was that which resulted in a distance of about 2.8 skin depths. We commented that we had also published results that showed a slightly different value for optimum frequency for other antenna configurations. We also saw that the optimum frequency was affected by the atmospheric noise, but that the attenuation v. frequency indicated that the optimum frequency was not a critical parameter. An adaptive communications system would allow us to make use of this parameter and so this would be a suitable topic for further investigation.

10.2 Antenna Design

In *chapter 3* we introduced the concept of specific aperture, and showed how this was a useful aid to antenna design, which we described in *chapter 4*. We demonstrated that a magnetic antenna system was the best for portable communications, but we noted that other antenna systems were also possible. As well as hybrid systems, such as an *E*-field transmitter and an *H*-field receiver, we noted that it was not necessarily straightforward to distinguish between cable-based systems such as r.f. single wire telephones (SWT) and long electric dipoles with earthed electrodes.

We suggested that there was much scope for further investigation in this area, including several novel or ‘exotic’ antennas such as the electric dipole toroid or ‘anapole’ and a rotating permanent magnet. Together with the use of amorphous metal tape as an alternative to ferrite, these present avenues for further investigation, which could lead to a significant improvement in the performance of sub-surface communications systems.

For conventional antennas the choice was essentially between an air-cored loop and a magnetic-cored solenoid. We saw that it was not possible to draw definite conclusions about the design of the transmitter and receiver antennas because the decisions such as tuned v. untuned, air-core v. ferrite, large v. small, wideband v. narrowband depended very much on the communications system being considered. This is more so than in ‘conventional’ radio; and also unlike true radio, we noted that with these near-field systems the type of antenna influenced the field propagation. However, the analysis in *chapters 4 to 6* can now be applied to specific systems – e.g. a radiolocation beacon, a pager system, a portable SSB speech system and a wideband a spread-spectrum sounding system – and conclusions drawn from these specific examples. This is seen as an essential next stage in this project.

We also concluded that it may be advantageous for a system may utilise different antennas for transmitting and receiving, and different antennas for the up-link and down-link paths.

10.3 Transmitter Design

We restricted our discussion of transmitter design in *chapter 5* to the power amplifier and the problems of driving an inductive load. We analysed the performance of tuned v. untuned antennas, and analogue v. digital drivers and a number of characteristics were tabulated. We saw that, although skin and proximity effects would affect the performance of a tuned antenna, there were circumstances where these effects were less significant. Importantly, we saw that, for maximum efficiency, a wideband transmitter should use a digital (class D) driver.

As with other aspects of the design, we saw that it was difficult to draw generalised conclusions. However, given specific design aims, the results of this chapter can be used to produce the ideal antenna configuration.

10.4 Noise and Receiver Design

In *chapter 6* we derived an expression for the signal to noise ratio (SNR) at the receiver. We showed how the atmospheric noise temperature F_a was related to the SNR, and that the high value of F_a at low frequencies was not reflected in a rise in the actual SNR. We described how the atmospheric noise would modify the optimum frequency, as derived in chapter 2, and we demonstrated a maximum useful size for the antenna. A significant conclusion was that, in the presence of typical levels of noise, the receiver antenna did not have to be very large.

We noted that the noise temperature data published by CCIR / ITU was obtained using a vertical electric dipole, and that this data could not be used to derive a figure for the magnetic equivalent noise field, thus highlighting the importance of collecting data using a magnetic field antenna.

The deduction of the maximum useful antenna size depended on the amplifier noise being minimal. We discussed this at length, analysing different antenna configurations of tuned and damped antennas and their noise-matching to the receiver amplifier. We discussed receiver amplifier design briefly, and concluded that the best choice of op-amp depended – as with many of our results – on the type of system being designed. Some op-amps have an extremely low noise power that allows them to be used with untuned antennas, but only at relatively low frequencies.

We outlined some schemes for the reduction of interference, including one we have devised, that makes use of the ‘quadrature phase space’ (QPS) behind an AM broadcast station. Mitigation of interference is a topic that would benefit from further investigation and we consider that the QPS method could, potentially, be the best way to improve the performance of narrowband induction radios using the LF band.

10.5 Channel Sounding

In *chapter 7* we proposed a method of channel sounding, and discussed the design of a wide-band sounder and data-logger. The design of this equipment has not been completed, and we were only able to present some preliminary results of background noise and interference in *chapter 9*. We consider the sounder to be a crucial part of an adaptive communications system but we may have to consider redesigning it to take account of the latest advances in digital

signal processing. It may be advantageous for the the data capture and signal averaging to be carried out by a DSP-based microprocessor, or field-programmable gate array rather than the hardware-based system we used.

In *chapter 8* we discussed the problem of system identification and synchronisation. We wished to use a binary sequence with a non-ideal auto-correlation function and so we developed a simple algorithm that allowed us to calculate the inverse, under correlation, of such a sequence. We described how such a sequence would be used in a code-locked synchronisation loop, that we intend to construct as the next stage of this project.

We omitted to present a design for the channel-sounder antennas. A preliminary analysis has been undertaken, but the precise details have not been finalised. Clearly this is an important part of the channel-sounder project, and the next stage of the development should be to design the transmitter and receiver antennas and verify that they will operate as desired with the wideband sounding sequence.

10.6 Concluding Remarks

In this thesis we have attempted to present a system level approach to the design of equipment for sub-surface ‘wireless’ communications, bringing together a study of the theory of electromagnetic wave propagation in a conducting medium and an analytical approach to antenna design. This has not been attempted before, and we hope that the material here will be of use for the design of sub-surface communications systems

In the absence of in-print reference material that could assist an engineer in the design of sub-surface communications systems, we feel that the further research and development of our communications system should take place alongside the production of suitable reference material, of which there is much that can be added to this thesis.

Appendices

A1 Author's Publications

♦ denotes a publication by invitation.

A1.1 Peer-reviewed & Published Conference Proceedings

- ♦ Gibson, D. & M. Darnell (1999), *Adaptive Digital Communications for Sub-Surface Radio Paths*, Proceedings of 5th International Symposium on Digital Signal Processing for Communications Systems (DSPCS '99), Perth–Australia, 237-244.
- Gibson, D. (1999), *Low-cost Implementation of a Speech Compression System*, Proceedings of 5th International Symposium on Digital Signal Processing for Communications Systems (DSPCS '99), Perth–Australia, 155-158.
- ♦ Gibson, D. & M. Darnell (2002), *A Channel Sounder for Low Frequency Sub-Surface Radio Paths*, Proceedings of 6th International Symposium on Digital Signal Processing for Communications Systems (DSPCS '02), Sydney: University of Wollongong, 49-54, ISBN 0 86418 791 2.

A1.2 Other Conference Presentations

- ♦ Gibson, D. (1994), *Induction Radio for Communications Through Rock*, Radio Solutions Conference Proceedings 1994, Low Power Radio Association (LPRA), www.lpra.org
- ♦ Gibson, D. (1997), *Using Low Power Radio for Communicating through the Ground*, Radio Solutions Conference Proceedings 1997, Low Power Radio Association (LPRA), www.lpra.org
- ♦ Gibson, D. (1998), *Sub-surface Radiolocation using Field Gradient Techniques*, URSI Symposium, Dublin.
- ♦ Gibson, D. (1999), *Sub-surface Radiolocation using Field Gradient Techniques*, URSI Symposium, York.
- Gibson, D. (2000), *Sub-Surface Radio Location at Skin-Depth Distances*, URSI Symposium, RAL Oxford.
- ♦ Gibson, D and M. Darnell (2000). *Simulation of an Underground Communications Antenna System*. MatLab DSP & Communications Conference, Reading: The Mathworks
- ♦ Gibson, D. (2001), *A Low Frequency Channel Sounder For Sub-Surface Communications*, Radio Solutions Conference Proceedings 2001, Low Power Radio Association (LPRA), www.lpra.org

A1.3 Peer-reviewed Papers

Gibson, D. (1996), 3-D Vector Processing of Magnetometer and Inclinometer Data, *Cave & Karst Science* 23(2), 71-76, ISSN 1356 191X

Gibson, D. (1996), How Accurate is Radiolocation?, *Cave & Karst Science* 23(2), 77-80, ISSN 1356 191X

A1.4 Papers in Preparation

Papers for publication are in preparation, based on chapters of this thesis, as follows

Chapter 3: *Specific Aperture: a Figure of Merit for Induction Loop Antennas*

Chapter 8: *Sequence Design for System Identification*

A1.5 Book Chapters

◆ Gibson, D. and M. Darnell (2002). A Channel Sounder for Low-Frequency Sub-Surface Radio Paths. In *Advanced Signal Processing for Communication Systems*. ed. T. Wysocki, M. Darnell and B. Honary. Dordrecht: Kluwer Academic Publishers. ISBN 1 4020 7202 3.

◆ † Gibson, D. (–). Communications in Caves. In *Encyclopaedia of Cave and Karst Science*. ed. J. Gunn. New York: Routledge.

◆ † Gibson, D. (–). Radiolocation. In *Encyclopaedia of Cave and Karst Science*. ed. J. Gunn. New York: Routledge.

† These essays have been accepted for publication in 2003

A1.6 CREG Articles

The author has written many articles for the Cave Radio & Electronics Group Journal. CREG is a special interest group of the British Cave Research Association. Many of the references quoted in this thesis have been included just for completeness, and for further reading for the dedicated reader.

A CD-ROM containing these articles may be available from the author at **12 Well House Drive, LEEDS, LS8 4BX. D.Gibson@caves.org.uk**. The author may also be reached via the British Cave Research Association at **enquiries@bcra.org.uk**.

A1.7 Magazine Articles

The author has many published magazine articles to his credit, and over 30 references in the Science Citation Index. However, only two are relevant to this thesis, namely...

Gibson., D. (1996), Designing an SSB Outphaser, *Electronics World* 102(1721) April 96, pp306-310 & 102(1722) May 1996, pp392-394

Gibson., D. (1991), Impedance Balancing Rule for Op-amps, *Electronic Product Design* 12(7) July 91, p17-19 (Re-written for *Electronics & Wireless World*, August 1993, p673).

A2 Propagation

A2.1 Calculating Fields using Retarded Potential

It is often convenient to represent the electromagnetic fields from a localised source as a potential function. Excepting the electrostatic potential, these functions may not have a clear physical significance but they do, nevertheless, provide a useful simplification in the analysis of many problems. There are several possible sets of potential functions; one commonly used is the *retarded potential*. It can be shown from Maxwell's equations – for example, see (Clemmow, 1973) §5.4 or (Ramo, Whinnery and Van Duzer, 1984) §3.19ff – that, for a current distribution \mathbf{J} , a vector \mathbf{A} exists such that

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \ddot{\mathbf{A}} = -\mu \mathbf{J} \quad \text{or} \quad (\nabla^2 + k^2) \mathbf{A} = -\mu \mathbf{J} \quad (\text{A2-1})$$

where a complex time-harmonic variation of $\exp(j\omega t)$ is assumed. The k term is the *wave number*, defined as

$$k = \frac{\omega}{v} = \frac{2\pi}{\lambda}, \quad \text{with} \quad v = \frac{1}{\sqrt{\mu\epsilon}} \quad (\text{A2-2})$$

and the remaining terms have their standard symbols¹. The electric and magnetic fields are defined in terms of \mathbf{A} , as

$$\mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A}, \quad \mathbf{E} = \frac{1}{j\omega\epsilon} \nabla \times \mathbf{H} \quad (\text{A2-3})$$

The reasoning that leads from Maxwell's equations to (A2-1) and from there to a solution for \mathbf{A} is well-rehearsed in, for example, (Clemmow, 1973), and we do not need to discuss it further here. Instead we will simply quote a solution to (A2-1) as being

$$\mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi} \int_{\tau} d\tau \frac{\mathbf{J}(\mathbf{r}') e^{-jkR}}{R} \quad (\text{A2-4})$$

where the integration is over the volume of the current distribution \mathbf{J} and R is the distance of the field point at \mathbf{r} from the volume element $d\tau$ at \mathbf{r}' . See *Figure A2-1*.

1 ϵ is permittivity, μ is permeability, v is the velocity of propagation, c is the speed of light, ω is angular frequency.

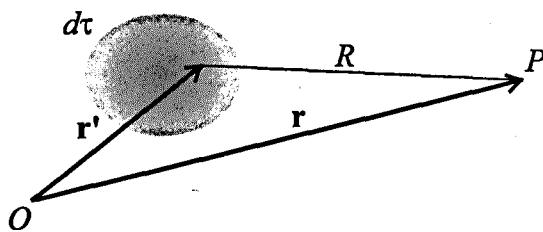


Figure A2-1 : Retarded potential.

The retarded potential $\mathbf{A}(\mathbf{r})$ at point P is calculated by integrating the current distribution $\mathbf{J}(\mathbf{r}')$ throughout region $d\tau$.

The current distribution is assumed to be within a linear, isotropic and homogeneous medium.

In a vacuum (for which we introduce a suffix: ϵ_0 , μ_0 and k_0) and if the field point is sufficiently distant that $|k_0(R - r)| \ll 1$, we can write

$$e^{-jk_0R} = e^{-jk_0(R-r)} e^{-jk_0r} \approx (1 + jk_0r - jk_0R) e^{-jk_0r} \quad (\text{A2-5})$$

and for the case of a loop of current in the ϕ direction, substituting the above into (A2-4) and replacing the volume integral with the appropriate loop integral gives

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} e^{-jk_0r} \left\{ (1 + jk_0r) \oint \frac{\mathbf{I} \cdot d\mathbf{l}}{R} - jk_0 \oint \mathbf{I} \cdot d\mathbf{l} \right\} \quad (\text{A2-6})$$

But, if the loop perimeter is much smaller than a wavelength, then the current can be assumed to be uniform and so we can take I outside the integrals. Noting that the second loop integral is then identically zero we are left with the expression

$$\mathbf{A}(\mathbf{r}) = I \frac{\mu_0}{4\pi} e^{-jk_0r} (1 + jk_0r) \oint \frac{\hat{\phi} \cdot d\mathbf{l}}{R} \quad (\text{A2-7})$$

and the evaluation of the loop integral leads directly to an expression for the vector potential of a small magnetic dipole.

For the case of an electric dipole, (A2-4) reduces to a line integral, which is easier to evaluate. We can consider two point charges $\pm q$ at the ends of a wire of length ℓ . We write the electrostatic dipole moment as $p = q\ell$. The charges are conveyed along the wire by a current $I = dq/dt$ so, in complex time-harmonic notation, we can write $p_d = I\ell = j\omega p$. If we assume that the field point is sufficiently far from this current distribution for it to look like a point source² then this is a simple current distribution which we can easily substitute into (A2-4). If the dipole is along the z -axis, then we obtain

$$\hat{\mathbf{A}} = \frac{\mu_0}{4\pi} p_d \frac{e^{-jk_0r}}{r} \quad (\text{A2-8})$$

² or, conversely, we understand that as ℓ contracts to zero, $q \rightarrow \infty$ such that p remains finite.

from which, with the aid of spherical polar co-ordinates and (A2-3), we can obtain the electric and magnetic fields from a small electric dipole. The magnetic dipole expression can then easily be obtained by applying a ‘duality’ substitution of

$$\mathbf{H} \rightarrow -\mathbf{E}, \quad \mathbf{E} \rightarrow \mathbf{H}, \quad \mu \leftrightarrow \epsilon, \quad p_d \leftrightarrow j\omega\mu m_d \quad (\text{A2-9})$$

where, $p_d = I \ell$ and $m_d = NI\pi a^2$ are the electric and magnetic dipole moments³.

This is the approach taken in many textbooks, where electric dipoles are usually considered to be more important in the context of antenna theory. However, should we wish to calculate the field from an extended current distribution, a more general evaluation of \mathbf{J} may be necessary, if a little tedious, and so it is helpful to consider a derivation obtained by solving (A2-7) directly.

The detail is shown in Appendix A2.2 and the result is that, for an infinitesimal circular loop, the vector potential – which is entirely in the ϕ direction – is

$$A_\phi = m_d \frac{\mu_0}{4\pi} e^{-jk_0 r} (1 + jk_0 r) \frac{\sin \theta}{r^2} \quad (\text{A2-10})$$

Using the spherical polar formulations for $\nabla \wedge \mathbf{A}$, equation (A2-10) leads us to the fields from an infinitesimal current loop of magnetic moment m_d ,

$$\mathbf{H} = \frac{m_d}{4\pi r^3} e^{-jk_0 r} \left\{ 2 \cos \theta (1 + jk_0 r) \hat{\mathbf{r}} + \sin \theta (1 + jk_0 r - (k_0 r)^2) \hat{\theta} \right\} \quad (\text{A2-11})$$

$$\mathbf{E} = \frac{j\omega\mu_0 m_d}{4\pi r^2} \sin \theta (1 + jk_0 r) e^{-jk_0 r} \hat{\phi} \quad (\text{A2-12})$$

A2.2 Vector Potential due to Small Circular Loop

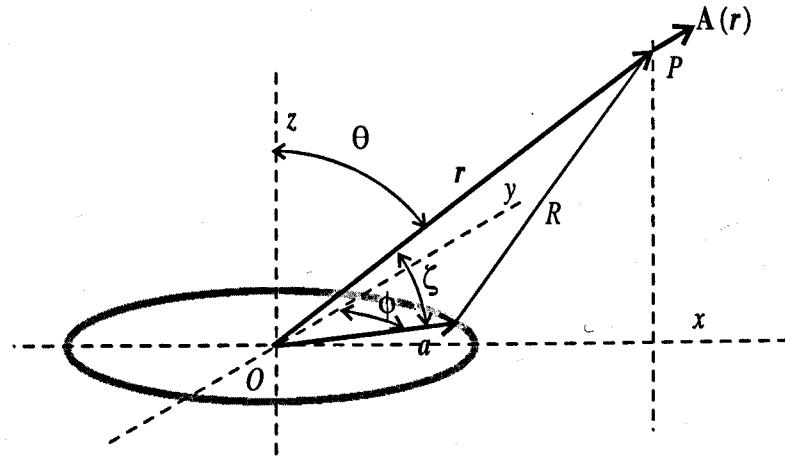


Figure A2-2 : Retarded potential due to a current loop.

3 p_d must not be confused with the electrostatic dipole moment $q\ell$.

We need to evaluate (A2-7) for the case of a loop of current. We will consider a small circular loop of radius a in the x - y or $\theta = \pi/2$ plane, centred at the origin. From arguments of symmetry we can deduce that \mathbf{A} will only have a ϕ component, and so it will be sufficient to consider a field point on the x axis, for which \mathbf{A} will be in the y direction. If ζ is the angle between the radius vector of the loop \mathbf{a} and the field point \mathbf{r} then the geometrical ‘law of cosines’ defines

$$R^2 = r^2 + a^2 - 2ar \cos \zeta \quad (\text{A2-13})$$

We need to find $1/R$ so we note that, if the field point is sufficiently distant that $a/r \ll 1$,

$$\frac{1}{R} = \frac{1}{\sqrt{r^2 + a^2 - 2ar \cos \zeta}} \approx \left(1 + \frac{a}{r} \cos \zeta \right) / r \quad (\text{A2-14})$$

Introducing spherical polar co-ordinates ϕ and θ , and writing the vectors \mathbf{a} and \mathbf{r} in Cartesian form $a(\cos \phi, \sin \phi, 0)$ and $r(\sin \theta, 0, \cos \theta)$ respectively, we note that their unit-vector dot-product will simply be

$$\cos \zeta = \cos \phi \sin \theta \quad (\text{A2-15})$$

Lastly, we note that the length element $d\ell$ can be written as $a d\phi$, and that, from symmetry, only the x component of this, $a d\phi \cos \phi$, will be significant. We can now substitute all the above terms into (A2-7) and obtain (A2-10).

A similar derivation to this, for a static magnetic dipole, appears in (Ramo *et al.*, 1984) §2.10.

A2.3 Exact Expression for Skin Depth

At high frequencies, or in a poor conductor, such that $\sigma/\omega\epsilon \ll 1$ we can derive an approximation to (2-9) of

$$k' = k_r - j/\delta, \quad \delta = \frac{2}{Z\sigma} \quad (\text{A2-16})$$

where Z is the wave impedance, defined as $\sqrt{\mu/\epsilon}$. We can derive an exact solution to (2-9) by applying De Moivre’s Theorem, as follows. We first write it as

$$k' = \left(\omega^2 \mu \epsilon - j \omega \mu \sigma \right)^{\frac{1}{2}} \quad (\text{A2-17})$$

and then choose suitable definitions of arbitrary parameters r and θ to write

$$k' = r(\cos 2\theta - j \sin 2\theta)^{\frac{1}{2}} \quad (\text{A2-18})$$

which, by De Moivre’s Theorem, we can write as

$$k' = r(\cos \theta - j \sin \theta) \quad (\text{A2-19})$$

From (A2-18) we can say that

$$r^2 \cos 2\theta = \omega^2 \mu \epsilon, \text{ and } r^2 \sin 2\theta = \mu \omega \sigma \Rightarrow r^2 = (\omega^2 \mu \epsilon) \sqrt{1 + \left(\frac{\sigma}{\omega \epsilon}\right)^2} \quad (\text{A2-20})$$

then a standard trigonometric relationship

$$1 - \cos 2\theta = 2 \sin^2 \theta \Rightarrow (r \sin \theta)^2 = \frac{1}{2} r^2 - \frac{1}{2} \omega^2 \mu \epsilon \quad (\text{A2-21})$$

allows us to write the attenuation coefficient as

$$\gamma = r \sin \theta = \sqrt{\frac{1}{2} \omega^2 \mu \epsilon \left(\sqrt{1 + \left(\frac{\sigma}{\omega \epsilon}\right)^2} - 1 \right)} \quad (\text{A2-22})$$

and the wave number as

$$k_r = r \cos \theta = \sqrt{\frac{1}{2} \omega^2 \mu \epsilon \left(\sqrt{1 + \left(\frac{\sigma}{\omega \epsilon}\right)^2} + 1 \right)} \quad (\text{A2-23})$$

A2.4 The Low-Frequency Window

Interestingly, we can demonstrate an optimum frequency without needing to invoke a magnetic induction effect. The magnitude of the co-planar ($\theta = \pi/2$) field strength can be written, from (2-18), as

$$H|_{\theta=\pi/2} = \frac{m_d}{4\pi r^3} e^{-T} \sqrt{T^2 (1+2T)^2 + (1+T)^2} \quad (\text{A2-24})$$

and the co-axial ($\theta = 0$) field as

$$H|_{\theta=0} = \frac{m_d}{4\pi r^3} 2e^{-T} \sqrt{T^2 + (1+T)^2} \quad (\text{A2-25})$$

Therefore, at a fixed distance from the transmitter, the field strength is a function of frequency and conductivity. This is shown in *Figure A2-3*, where the x -axis is the parameter T , which can be considered as a normalised frequency.

The coplanar field shows an unexpected rise in field strength with T , the maximum occurring at $T \approx 1.62$ for which the T term has risen above its quasi-static ($T = 0$) value of unity to 1.45 – a rise of 3.3dB. The interpretation is that, for a given distance from the transmitter, the coplanar field will be a maximum if we arrange the frequency such that the transmission distance is about 1.6 skin depths.

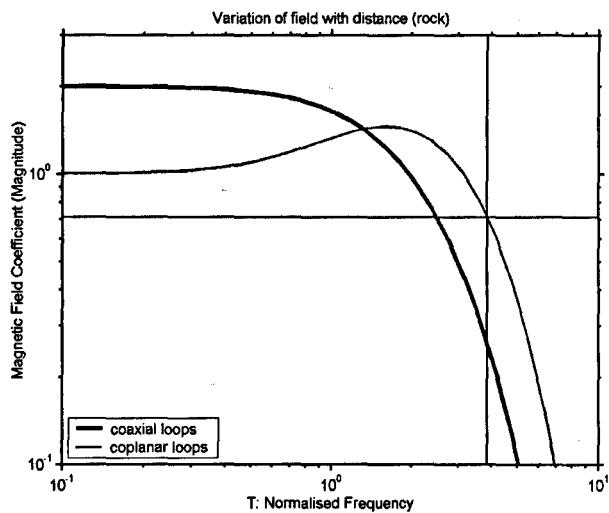


Figure A2-3 : Low-frequency window.

A low-frequency 'window' is evident for the coplanar orientation. The 'cross-hairs' indicate the -3dB point for the co-planar orientation, at $T = 3.85$.

This effect is well-known, and is sometimes referred to as the *low-frequency window*⁴. It is analysed by (Gabillard, Degauque and Wait, 1971), who use this to characterise the medium, defining a -3dB point at $T_c = 3.85$ where the field strength at a given distance is $1/\sqrt{2}$ times its quasi-static value. To make this clearer, we can re-write the x -axis of Figure A2-3 in terms of frequency, noting that

$$T = \frac{r}{\delta} \Rightarrow \omega = 2 \frac{T^2}{\mu\sigma} \frac{1}{r^2} \quad (2-26)$$

and so we can relate the 'cut-off' frequency⁵, ω_c , to a distance, r_c , by

$$\omega_c = 2 \frac{T_c^2}{\mu\sigma} \frac{1}{r_c^2}, \quad \text{with } T_c = 3.85 \quad (2-27)$$

For example, at a distance of 100m, and with a conductivity of 0.01S/m, we would have an available bandwidth of 37.5kHz. At 1000m the bandwidth would be 375Hz. In seawater ($\sigma = 4/\text{m}$) the bandwidth would be 23Hz over 200m or 2.3kHz over 20m.

A similar analysis is possible for the co-axial arrangement, which demonstrates a lower bandwidth, see *Table A2-1* below.

- 4 A full analysis, taking into account the permittivity, additionally demonstrates a high-frequency window but, since this usually occurs at GHz or optical frequencies we need not consider it here. Permittivity tends to a value of ϵ_0 at sufficiently high frequencies and conductivity also ceases to obey a simple law, but we can note that a high-frequency window clearly exists in sea-water at optical frequencies.
- 5 Cut-off and characteristic frequency (2-16) share the same symbol and must not be confused.

co-axial loops:	-3dB at $T_c = 1.33$	Peak 2.00 at $T = 0$
co-planar loops:	-3dB at $T_c = 3.85$	Peak 1.45 at $T = 1.62$

Table A2-1 : Bandwidth of low frequency window.
Compare this with Table 2-5.

In his review of communications in South African mines, (Austin, 1987) appears to misinterpret the (Gabillard *et al.*, 1971) paper, confusing the low frequency window with the characteristic frequency, and additionally deriving an erroneous expression for the *range* of communication as a function of characteristic frequency, bandwidth and skin depth. Austin observes, not surprisingly, that the equations did not agree with experimental results to within an order of magnitude. In (Wait, 1992) it is noted that as the permittivity of the medium increases (causing the ‘good conductor’ approximation to become less valid), the peak of the window becomes more pronounced.

A2.5 Computer Simulation of a Buried VMD

A2.5.1 MatLab Simulation

The expressions for the horizontal and vertical fields from a buried horizontal loop, or vertical magnetic dipole (VMD) can be solved using the aid of a computer program such as MatLab, which is also a convenient tool for presenting the results graphically.

The programs are presented on the bound-in CD-ROM, and indexed in appendix A9. Only a few salient points will be discussed now⁶. The functions **PQintegral** and **sommerfeld** perform the task of evaluating the expression

$$(P, Q) = \int_0^\infty dx \frac{x^3 e^{-xZ}}{x+U} e^{-U} (J_1(xD), J_0(xD)), \quad \text{with } U^2 = x^2 + jT^2 \quad (\text{A2-28})$$

The integrand is evaluated for a particular value of the dummy variable λ (represented by x in the programs) by the function **sommerfeld** (Table A2-2, below) and the result is numerically integrated by the function **PQintegral** (Table A2-3, below) using an adaptive Lobatto quadrature method that is built in to MatLab’s function **quadl**.

The function **PQintegral** takes four arguments – values for the three normalised parameters D , T and Z , and an integer representing the required field evaluation (uplink P , downlink P or Q). **sommerfeld** takes an additional argument for the value of the dummy variable x .

6 The reader will need to be familiar with MatLab in order to fully understand the programs listed here. MatLab uses a conventional ‘third-generation’ algorithmic language, but with several features that distinguish it from similar languages. The most important of these is that program variables are treated as matrices, and there is therefore a distinction between, for example, vector or matrix multiplication and element-by-element multiplication of those entities.

```

function a = sommerfeld(x,D,T,Z,mode)
% (x,D,T,Z,mode) returns integrand of a Sommerfield integral,
% used for calculations of Wait's magnetic field equations.
%
% x is the dummy variable used by the integration routine
% See PQintegral for further information.

u = sqrt( x.*x + (T.^2)*j );
if T==0,                                     % special case
    a = (x) .* exp(-u) .* exp(-x.*Z) /2;
else
    a = (x.^2) .* exp(-u) .* exp(-x.*Z) ./(x + u);
end
if mode==1                                     % uplink P field
    a = x .* besselj(1, x.*D) .* a;
elseif mode==2                                  % downlink P field
    a = u .* besselj(1, x.*D) .* a;
elseif mode==3                                  % Q field
    a = x .* besselj(0, x.*D) .* a;
else
    a=0;
end

```

Table A2-2 : MatLab program listing for function 'sommerfeld'.
The T parameter in this program is "Wait's" definition, notated by \mathcal{T} in the text.

```

function a = PQintegral(D,T,Z,mode)
% PQintegral(D,T,Z,mode) performs a Sommerfeld integral
% mode=1: returns P(D,T,Z) for up-link direction
% mode=2: returns P(D,T,Z) for down-link direction
% mode=3: returns Q(D,T,Z) for either direction

% Z is normalised height of observation point
% D is normalised offset of observation point
% T is normalised depth of transmitter

TOL = 1E-8; % the default is 1E-6
TRACE = 0;
XLIMIT = 30;

a = quadl(@sommerfeld,0,XLIMIT,TOL,TRACE,D,T,Z,mode);

% Numerically evaluate integral, adaptive Lobatto quadrature.

```

Table A2-3 : MatLab program listing for function 'PQintegral'.
The T parameter in this program is "Wait's" definition, notated by \mathcal{T} in the text.

Example. To find the Q field on the surface, at an offset of $D = \sqrt{2}$ from GZ, an elevation of $Z = 0$, and with a ground conductivity of zero ($T = 0$) a keyboard of,

PQintegral(sqrt(2), 0, 0, 3)

results in the displayed answer of **-1.0984e-012** confirming what we know from a geometrical understanding of the field, that the field line at this offset is horizontal. Similarly,

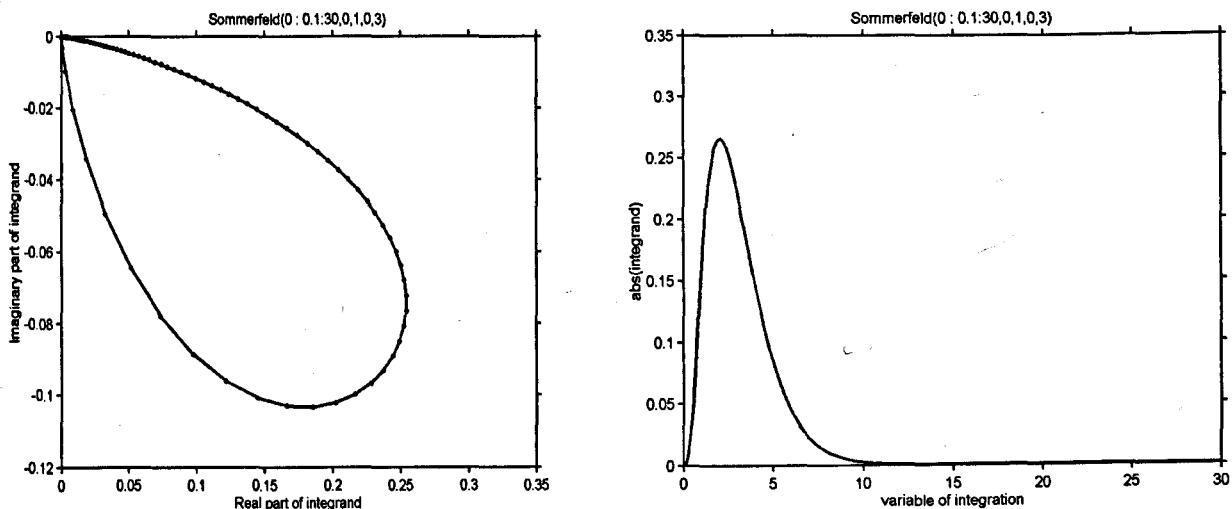
PQintegral(0, 0, 0, 3)

results in **1.0000** as expected. However, with a finite conductivity, say $T = 0.5$, the answers we obtain are **-0.0093 - 0.0168i** and **0.9860 - 0.0796i** respectively. We can see that the null in the field has been partially filled in (it is now -34dB) and that both fields demonstrate a phase delay. The change in phase is particularly significant to a discussion of radiolocation.

A2.5.2 Inspection of the Sommerfeld Integrand

The next stage in the analysis will be to present this data graphically, but first we need to confirm the integration parameters used in Matlab's **quad1** function. The integrand of (A2-28) is plotted in *Figure A2-4ff*. By inspecting these and similar plots we were able to deduce that an upper limit of 30 and a tolerance of 1E-8 were usually sufficient for the numerical integration. The multiple peaks in *Figure A2-5b* etc. are due to the presence of the Bessel function.

We observed that, for large values of the normalised offset D , the integrand did not completely settle down by $x = 30$, and that it oscillated many times. By comparing the results with what we would expect for the $T = 0$ case, we deduced that the integration tolerance needed to be extended to 1E-10, and that the limit of integration should be extended to 50. Our observations are similar to those reported by (King and Sandler, 1977) who were considering the propagation at 100kHz in lake water over a distance of 20km. In that situation, the oscillations of the integrand, introduced by the Bessel function, meant that a numeric integration was difficult. The graphs presented in chapter 2 all use the routine of *Table A2-3* and so, should we wish to investigate secondary fields at high offsets, we must remember to check the validity of the numeric integration.



*Figure A2-4 : Integrand of Sommerfeld integral.
Argand diagram (left) and magnitude (right). Parameters D=0, T=1, Z=0; mode: Q-field. Integrand evaluated for x=0 to 30 in steps of 0.1.*

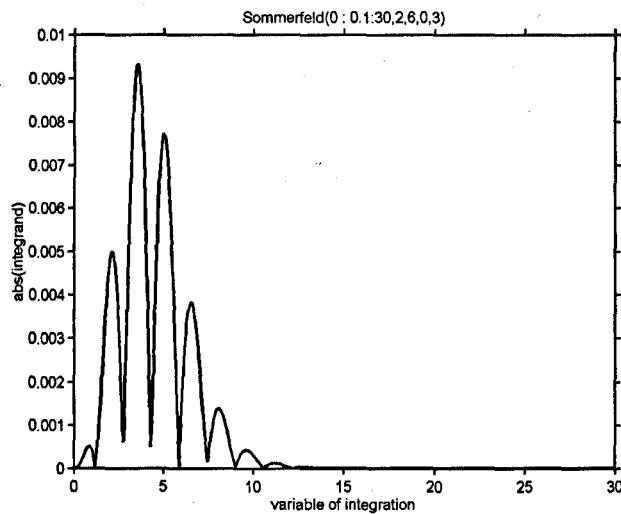
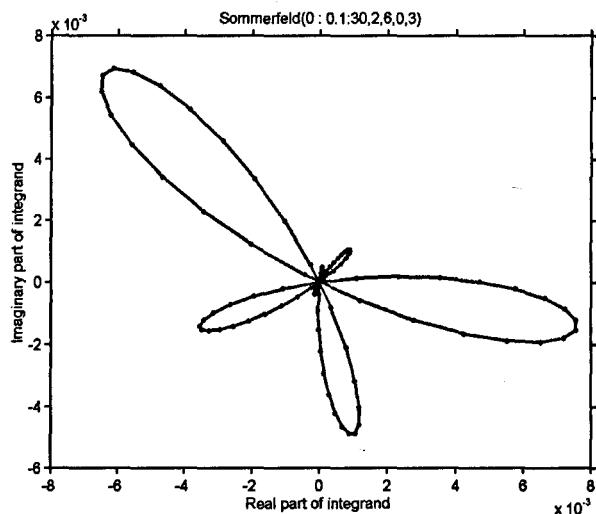


Figure A2-5 : as Figure A2-4 with parameters as given.
 $D=2, T=6, Z=0$; mode: Q-field

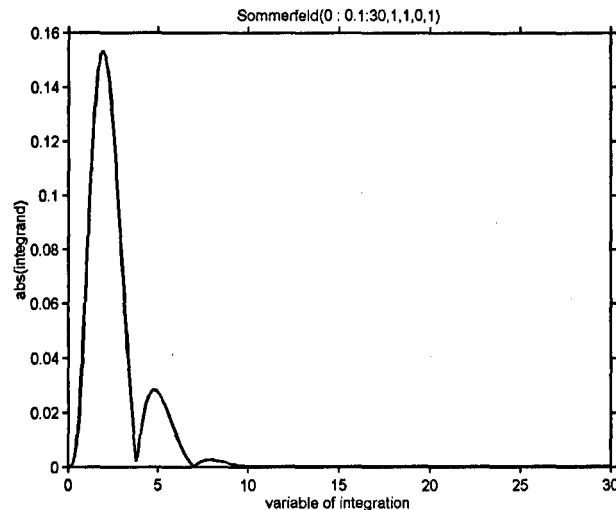
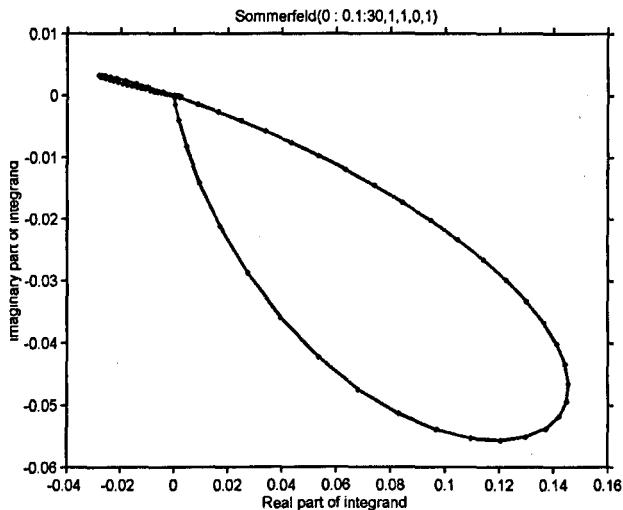


Figure A2-6 : as Figure A2-4 with parameters as given.
 $D=1, T=1, Z=0$; mode: up-link P-field

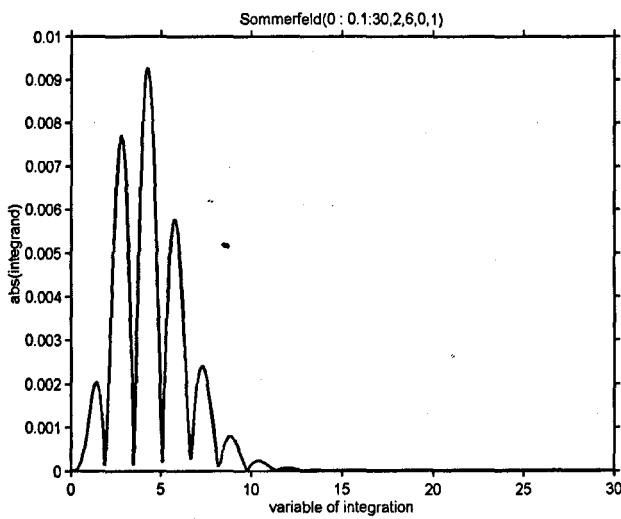
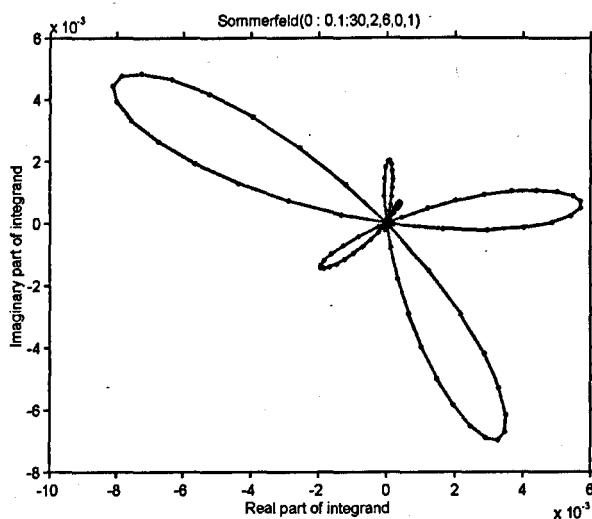


Figure A2-7 : as Figure A2-4 with parameters as given.
 $D=2, T=6, Z=0$; mode: up-link P-field

A2.6 References

- Austin, B. A. (1987). Radio Communication through Rock. *Electronics & Wireless World* **93**(1619), 943-946.
- Clemmow, P. C. (1973). *An Introduction to Electromagnetic Theory*. Cambridge: Cambridge University Press.
- Gabillard, P. D., P. Degauque and J. R. Wait (1971). Subsurface Electromagnetic Telecommunication - a Review. *IEEE Transactions on Communication Technology COM-19*(6), 1217-1228.
- King, R. and B. Sandler (1977). Subsurface Communication between Dipoles in General Media. *IEEE Transactions on Antennas and Propagation AP-25*(6), 770-775.
- Ramo, S., J. R. Whinnery and T. Van Duzer (1984). *Fields and Waves in Communication Electronics*. New York: John Wiley. (2nd Edition).
- Wait, J. R. (1992). Antennas in the Geophysical Environment - Some Examples. *IEEE Proceedings* **80**(1), 200-203.

A3 Driving an Inductive Load

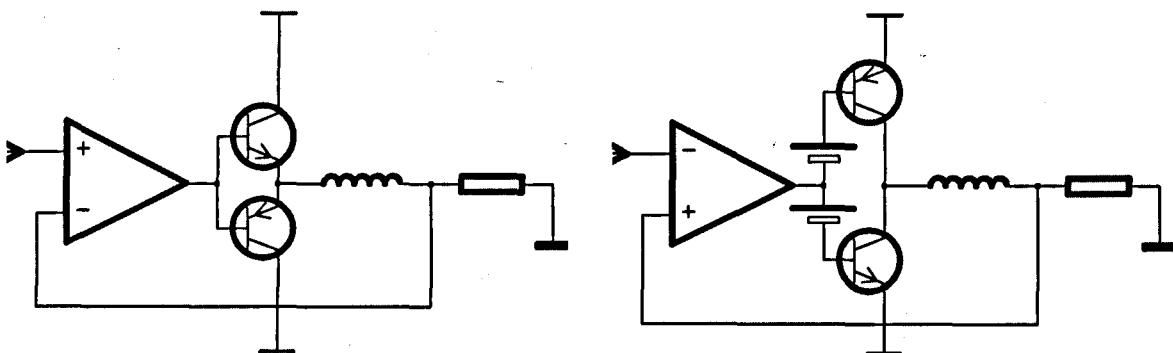


Figure A3-1 : Class-B/C amplifier driving an inductive load
Left: push-pull configuration. Right: Complementary configuration.

Consider the amplifiers shown, in outline form, in **Figure A3-1** and discussed in §5.2.1. The purpose of this analysis is to find the power dissipation in the output transistors and the load resistor when the voltage and current are not in phase.

With this circuit, the ‘top’ transistor supplies positive-going current to the load and the ‘bottom’ transistor sinks negative-going current. The voltage across the load will lead the current by a phase angle ϕ and will go negative whilst the current is still positive. This means that although the output of the amplifier is negative, current is still being drawn through the top transistor, which therefore has to drop a larger voltage than if the current were being supplied from the bottom transistor. Similarly, when the output voltage is zero, the current will be finite, leading to dissipation in the output transistors that would not be the case for a resistive load. It is these effects that lead to the inefficiency when driving a reactive load.

With feedback we can presume that the high gain of the pre-amp stage allows the power transistors to be biased at the level necessary to achieve the demanded output, which simplifies the analysis. Without feedback, the operation is similar, but not as easy to justify. In all cases we have verified the operation by running a circuit simulation using PSPICE.

The push-pull output stage is basically a voltage follower, but the output node can be ‘forced’ by the output current. For example, if the base voltages are positive, a negative current flow (i.e. *into* the emitters of the output transistors, can force the emitter nodes to rise, cutting off the top transistor and turning on the bottom one. The result of this is that, regardless of the demand voltage at the input to the PA, the top transistor is on when the load current is positive, and the bottom transistor is on when the load current is negative. The complementary output stage, on the other hand, is a current follower and, in an open-loop system, it is the output *current* that follows the input voltage demand, with the output voltage effectively ‘floating’ to the necessary level.

A3.1 Sinusoidal Signals

We begin by noting a couple of results that will be useful later on in the analysis. Firstly, in a series L-R circuit the phase angle ϕ of the voltage, with respect to the current, is defined by its tangent, but we can also define the **power factor**, $\cos \phi$, as follows.

$$\tan \phi = \frac{\omega L}{R} = Q \Rightarrow \cos \phi = \frac{1}{\sqrt{1+Q^2}} \quad (\text{A3-1})$$

and secondly, the magnitude of the impedance, $|Z|$ can be defined as

$$|Z| = |R + j\omega L| = R\sqrt{1 + \tan^2 \phi} = \frac{R}{\cos \phi} \quad (\text{A3-2})$$

A3.1.1 Power Transfer Efficiency

We suppose that the amplifier is controlling a sinusoidal current in the sense resistor (which may be a physical resistor, or the resistive component of the induction loop load). The voltage waveform across the load will lead the current by an amount ϕ , depending on the ratio of resistance to inductance. Let the current waveform in the load be $I \sin \theta$ – the voltage will therefore be $U_l \sin(\theta + \phi)$ where θ is a short-hand for ωt . We consider the positive half-cycle of current, which is sourced by the top transistor. The instantaneous power drawn from the amplifier, with rail voltage U_p is simply

$$(U_p - U_l \sin(\theta + \phi)) \cdot I \sin \theta \quad (\text{A3-3})$$

and, over a half-cycle, the mean power dissipation, P_p is found from

$$\begin{aligned} P_p &= I \frac{1}{\pi} \int_0^\pi d\theta (U_p - U_l \sin(\theta + \phi)) \cdot I \sin \theta \\ &= I \left(\frac{2}{\pi} U_p - \frac{1}{2} U_l \cos \phi \right) \end{aligned} \quad (\text{A3-4})$$

The mean power dissipated in the load is simply

$$P_l = \frac{I^2 R}{\pi} \int_0^\pi \sin^2 \theta d\theta = \frac{1}{2} I^2 R \quad (\text{A3-5})$$

so the ratio of the load power to the amplifier power dissipation, using (A3-2), is

$$\begin{aligned} \frac{P_l}{P_p} &= \frac{\frac{1}{2} I R}{\frac{1}{\pi} 2 U_p - \frac{1}{2} U_l \cos \phi} = \frac{1}{\frac{4}{\pi} \frac{U_p}{U_l \cos \phi} - 1} \\ &\Rightarrow \frac{P_l}{P_l + P_p} = \frac{\pi}{4} \frac{U_l}{U_p} \cos \phi \end{aligned} \quad (\text{A3-6})$$

It is easy to eliminate $\cos \phi$ from this by using (A3-1) but we also need to eliminate U_l . We can write U_l as a fraction of U_p but it is more useful to write it as a fraction of the maximum

possible U_l – that is the maximum output before clipping occurs. If we call this fraction the ‘gain’ g we have

$$\frac{U_l}{U_p} = g \frac{U_{l,max}}{U_p} \quad (\text{A3-7})$$

but we know that the right-hand side of this must be

$$\frac{U_{l,max}}{U_p} = \frac{\sqrt{R_l^2 + (\omega L)^2}}{\sqrt{(R_l + R_p)^2 + (\omega L)^2}} = \frac{\sqrt{1+Q^2}}{\sqrt{\left(1 + \frac{R_p}{R_l}\right)^2 + Q^2}} \quad (\text{A3-8})$$

where R_l is the load resistance and R_p is the collector-emitter resistance of the output transistor, assuming that it has a zero saturation voltage. (This particular representation may be more suited to MosFet transistors, where R_p represents the ‘ $R_{DS, on}$ ’ figure of the device).

We can simplify the expression by defining a matching coefficient as

$$\eta = \sqrt{\frac{R_l}{R_p}} \quad (\text{A3-9})$$

The result of these manipulations is that the power transfer efficiency – the load power divided by the total power consumed – is

$$\frac{P_l}{P_l + P_p} = \frac{\pi}{4} g \frac{\eta}{m} \quad (\text{A3-10})$$

with

$$m = \sqrt{(\eta + 1/\eta)^2 + \eta^2 Q^2} \quad (\text{A3-11})$$

A3.1.1.1 Observation on matching co-efficient

It is important to realise that the matching criterion, (A3-9), relates the load resistance to the ‘saturation resistance’ of the output devices. It does not relate it to the ‘output impedance’ of the amplifier. If that latter term has any meaning, it is probably the output impedance of a Thévenin-equivalent circuit model. However, since the amplifier has feedback, the figure would be very close to zero (for voltage feedback) or very high (for current feedback). This impedance would be relevant to transmission line matching, but it is not the figure we need to use for power-transfer matching.

If, instead of a ‘saturation resistance’ R_p , it is more appropriate to consider a saturation voltage then a different analysis must be undertaken.

A3.1.2 Load Power Dissipation

We also need to know the load power in absolute terms, and so, from (A3-5), using (A3-2) we write

$$P_l = \frac{1}{2} I^2 R = \frac{1}{2} \left(\frac{U_l \cos \phi}{R} \right)^2 R = \frac{1}{2} \frac{U_p^2}{R} \left(g \frac{U_{l,\max}}{U_p} \cos \phi \right)^2 \quad (\text{A3-12})$$

to which we apply the same substitutions as before, to obtain

$$P_l = \frac{1}{2} g^2 \frac{U_p^2}{R_l} \frac{\eta}{m} = \frac{1}{2} g^2 \frac{U_p^2}{R_p} \frac{1}{m} \quad (\text{A3-13})$$

The presence of the m term demonstrates the principle of matching for optimum power transfer. The load power is a maximum when m is a minimum, and – for a resistive load with $Q = 0$ – this is when η , the matching coefficient is unity.

Alternatively, substituting directly for $\cos \phi$ in (A3-12) leads to

$$P_l = \frac{1}{2} \frac{U_l^2}{R_l} \frac{1}{1+Q^2} \quad (\text{A3-14})$$

A3.1.3 Amplifier Power Dissipation

Starting from (A3-4) we can derive

$$\Rightarrow P_p = \frac{U_p^2}{R_l} \frac{1}{2} \frac{g\eta}{m} \left(\frac{4}{\pi} - \frac{g\eta}{m} \right) \quad (\text{A3-15})$$

which also demonstrates a ‘matching’ condition because a term of the form $x(k-x)$ is a maximum when $x = \frac{1}{2}k$. This result is well-known for an audio amplifier driving a resistive load – for which $\eta/m \approx 1$ – where it shows that for a voltage output that is $2/\pi$ of the rail voltage, the amplifier dissipates the least power.

A3.2 Square-Wave Signals

A3.2.1 Complementary and H-Bridge Output

For a digital output from the amplifier we can simplify the analysis by assuming a) a square wave with a 1:1 duty cycle, b) that $Q \gg 1$, so that the current waveform in the load is a triangle¹ and c) that we operate without feedback. A complementary configuration – whether it be the bipolar circuit shown above, or a mosfet H-bridge – exhibits the characteristics that

1 The requirement could be stated as $L/R \gg T/2$ – that is, the L-R time constant is long compared with the rise-time of the waveform. This equates to $Q \gg \pi$.

- a) rising current flows from the top transistor,

(note that this is not the same condition we observed earlier for sine-wave signals).

- b) the output voltage follows the input voltage, and
 c) positive voltage causes rising current.

Thus, with the top transistor fully on, the current in the load rises from $-I_{max}$ to $+I_{max}$ at a rate $I/t = U_p/L$. To find the load power dissipation we simply integrate this current, as follows.

$$\begin{aligned} P_l &= R_l \frac{2}{T} \int_{-T/4}^{+T/4} \left(\frac{U_p}{L} \right)^2 t^2 dt = R_l \frac{2}{T} \left(\frac{U_p}{L} \right)^2 \left[\frac{1}{3} t^3 \right]_{-T/4}^{+T/4} \\ &= \frac{U_p^2}{R_l} \frac{\pi^2}{12} \frac{1}{Q^2} \end{aligned} \quad (\text{A3-16})$$

The amplifier power dissipation is R_p / R_l times the above expression, and so the ratio of the two, from which we can obtain the efficiency, is

$$\frac{P_l}{P_p} = \frac{R_l}{R_p} = \eta^2 \Rightarrow \frac{P_l}{P_l + P_p} = \frac{1}{1 + 1/\eta^2} \quad (\text{A3-17})$$

This analysis applies to complementary and H-bridge drivers operated with 1:1 duty cycles. Clearly we can adapt the expression to allow for different duty cycles, for pulse-width modulation and – for an H-bridge – a re-circulation time. However, we can surmise that the basic results will be similar, namely that the load power is proportional to $1/Q^2$ and the efficiency is a function of the matching coefficient.

A3.2.2 Push-Pull Output

The push-pull output stage demonstrates slightly different behaviour due to the ‘node forcing’ we discussed earlier, the difference being that it is *positive* current and not *rising* current that flows in the top transistor. Additionally, because the device is a voltage follower we have to know the demanded load voltage in order to analyse the circuit, therefore it helps the analysis to assume that we provide voltage feedback. Whilst the top transistor is on, the amplifier power dissipation is

$$\begin{aligned} P_p &= \frac{4}{T} \left\{ \int_0^{T/4} dt (U_p - U_l) \cdot \frac{U_l t}{L} + \int_0^{T/4} dt (U_p - U_l) \cdot \frac{U_l t}{L} \right\} \\ &= \frac{4}{T} \cdot \left(\frac{U_l}{L} \right) \cdot \left[\frac{1}{2} t^2 \right]_0^{T/4} \{ 2U_p \} \\ &= \frac{U_l}{U_p} \frac{U_p^2}{R_l} \frac{\pi}{2} \frac{1}{Q} \end{aligned} \quad (\text{A3-18})$$

which, if we assume $\eta \geq 1$ and $Q \gg 1$, approximates to

$$P_p = g \frac{U_p^2}{R_l} \frac{\pi}{2} \frac{1}{Q} \quad (\text{A3-19})$$

The load power, given by a similar expression to (A3-16), is

$$P_l = g \frac{U_l^2}{R_l} \frac{\pi^2}{12} \frac{1}{Q^2} \quad (\text{A3-20})$$

so the efficiency is

$$\frac{P_l}{P_p} = g \frac{\pi}{6} \frac{1}{Q} \quad (\text{A3-21})$$

which is distinctly lower than the value for the complementary output stage.

A3.3 Partially-tuned Load

For an untuned load, we substituted the expression for $\cos \phi$ (A3-1) into (A3-6) and proceeded to derive the expressions we require. We can take a similar approach to deal with the case of a partially resonant load. The impedance of a series *LCR* resonant circuit is

$$\frac{Z}{R} = 1 + \frac{j\omega L}{R} + \frac{1}{j\omega CR} = 1 + jQ_0 \overline{\Omega} \quad (\text{A3-22})$$

where we use Q_0 to indicate that the Q -factor is calculated at ω_0 , and we define

$$\overline{\Omega} = \Omega - \frac{1}{\Omega}, \quad \text{with} \quad \Omega = \frac{\omega}{\omega_0} \quad (\text{A3-23})$$

From this we can see that

$$\cos \phi = \frac{1}{\sqrt{1 + Q_0^2 \overline{\Omega}^2}} \quad (\text{A3-24})$$

and, using this notation, this replaces (A3-1), which we could re-define as

$$\cos \phi = \frac{1}{\sqrt{1 + Q_0^2 \Omega^2}} \quad (\text{A3-25})$$

(i.e. it contains the term Ω instead of $\overline{\Omega}$). Therefore, the alteration to the previously-defined expressions for an untuned, inductive load is simply to replace Q by $Q_0 \overline{\Omega}$.

A4 Receiver Noise-Matching

A4.1 Noise Effects of Bias Resistor

If a bias resistor is connected to the input of the amplifier then two additional sources of noise are present – the thermal noise of the resistor, and a contribution from the noise current of the amplifier. The noise factor, given in equation (6-3) becomes

$$F_n = 1 + F_a \frac{R_r}{R} + \frac{1}{4kT} \left(\frac{\hat{v}_n^2}{R\eta^2} + \hat{i}_n^2 R\eta^2 + \frac{\hat{i}_n^2 R_b^2}{R\eta^2} \right) + \frac{R_b}{R\eta^2} \quad (\text{A4-1})$$

If R_b is low in value, with respect to the apparent source resistance $R\eta^2$, then the last term can be ignored, leaving just the effect of the noise current. We can incorporate this in the noise-matching term by writing

$$\frac{P_n}{4kT} \frac{R'_n}{R_n} \left(\frac{R'_n}{R\eta^2} + \frac{R\eta^2}{R'_n} \right) \quad (\text{A4-2})$$

with $R'_n = \sqrt{R_n^2 + R_b^2} = R_n \sqrt{1 + \left(\frac{R_b}{R_n} \right)^2}$ (A4-3)

showing that R_b will have a negligible effect if it is low in value with respect to the noise resistance R_n , i.e. the ratio of the noise voltage spectral density to the noise current spectral density. Including the effects of R_b only serves to cloud the issues at this stage of the analysis, so we will ignore it.

A4.2 Noise Factor of Parallel-tuned Antenna

A4.2.1 Notation

In addition to the notation described in §6.3 it is convenient to define a ‘gain’ term Γ , which is the output voltage (across the C) in terms of the input voltage (induced in the loop). Thus,

$$\Gamma = \frac{1/j\omega C}{1/j\omega C + j\omega L + R + R_d} = \frac{-jQ_0/\Omega}{jQ_0\bar{\Omega} + \theta} \quad (\text{A4-4})$$

with $\theta = 1 + \frac{R_d}{R}$ (A4-5)

Note that Q_0 is the quality factor *without* the damping resistor R_d present; that is

$$Q_0 = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 C R} \quad (\text{A4-6})$$

The impedance, looking in to the loop from the output is

$$\begin{aligned} Z &= (1/j\omega C) \parallel (j\omega L + R + R_d) \\ &= \Gamma (j\omega L + R + R_d) \\ &= \Gamma R (jQ_0 \Omega + \theta) \end{aligned} \quad (\text{A4-7})$$

A4.2.2 3dB Bandwidth

In subsequent calculations we will need to know the 3dB bandwidth, which is found by considering

$$|\Gamma|^2 = \frac{(Q_0/\Omega)^2}{(Q_0\bar{\Omega})^2 + \theta^2} \quad (\text{A4-8})$$

Initially, we will suppose that Ω will not deviate much from unity over the bandwidth, and so the 3dB points of the above equation will be at

$$Q_0 \bar{\Omega} = \theta \quad (\text{A4-9})$$

If this is so, then we can derive, from this expression, the condition

$$\Omega = 1 \pm \frac{\theta}{2Q_0} \quad (\text{A4-10})$$

So, with the condition $Q_0 \gg \theta$, our original supposition is correct, and the 3dB bandwidth will be given by (A4-9).

A4.2.3 General Expression

The general expression for the noise factor of a tuned antenna – whether at resonance or not – is found by combining

$$\text{a) Signal voltage} \quad \omega \mu H \cdot \Phi \sqrt{R} \cdot \eta |\Gamma| \quad (\text{A4-11a})$$

$$\text{b) Thermal noise} \quad \sqrt{4kT_0 B(R + R_d)} \cdot \eta |\Gamma| \quad (\text{A4-11b})$$

$$\text{c) Atmospheric noise} \quad \sqrt{4kT_0 B F_a R_r} \cdot \eta |\Gamma| \quad (\text{A4-11c})$$

$$\text{d) Amplifier noise voltage} \quad \hat{v}_n \sqrt{B} = \sqrt{\hat{P}_n R_n} \sqrt{B} \quad (\text{A4-11d})$$

$$\text{e) Amplifier noise current in } Z \quad \hat{i}_n \eta^2 |Z| \sqrt{B} = \sqrt{\frac{\hat{P}_n}{R_n}} \eta^2 |Z| \sqrt{B} \quad (\text{A4-11e})$$

and we will divide each of these terms by $\sqrt{4kT_0 B R} \cdot \eta |\Gamma|$ to get

$$\text{a) Signal voltage} \quad \omega \mu H \cdot \Phi / \sqrt{4KTB} \quad (\text{A4-12a})$$

b) Thermal noise $\sqrt{\theta}$ (A4-11b)

c) Atmospheric noise $\sqrt{F_a R_r / R} = F_k \Phi$ (A4-11c)

d) Amplifier noise voltage $\sqrt{\frac{\hat{P}_n}{4kT} \frac{R_n}{R \cdot \eta^2 |\Gamma|^2}}$ (A4-11d)

e) Amplifier noise current in Z $\sqrt{\frac{\hat{P}_n}{4kT} \frac{\eta^2 \Gamma R (jQ_0 \Omega + \theta)^2}{R_n \cdot R \cdot \eta^2 |\Gamma|^2}}$ (A4-11e)

The output SNR is

$$\frac{S}{N} = \frac{\omega \mu H}{\sqrt{4kTB}} \Phi / \sqrt{F_n} \quad (\text{A4-13})$$

with $F_n = \theta + (F_k \Phi)^2 + \frac{\hat{P}_n}{4kT} A_n$ (A4-14)

with the A_n term given by

$$\begin{aligned} A_n &= \frac{R_n}{R \cdot \eta^2 |\Gamma|^2} + \frac{\eta^2 R ((Q_0 \Omega)^2 + \theta^2)}{R_n} \\ &= \frac{R_n}{R \cdot \eta^2} \frac{\Omega^2}{Q_0^2} (\theta^2 + (Q_0 \bar{\Omega})^2) + \frac{R \cdot \eta^2}{R_n} Q_0^2 \Omega^2 (1 + (\theta/Q_0 \Omega)^2) \\ &= \left\| \frac{R_n}{R \cdot \eta^2 Q_0^2} \frac{\sqrt{\theta^2 + (Q_0 \bar{\Omega})^2}}{\sqrt{1 + (\theta/Q_0 \Omega)^2}} \right\| \Omega^2 \sqrt{\theta^2 + (Q_0 \bar{\Omega})^2} \sqrt{1 + (\theta/Q_0 \Omega)^2} \end{aligned} \quad (\text{A4-15})$$

A4.2.4 Approximations

A4.2.4.1 At Resonance

At resonance, we have $\Omega = 1$ and $\bar{\Omega} = 0$. If, in addition, there is no damping resistor (so $\theta = 1$) and we have $Q_0 \gg 1$, then we obtain the expression given in the main text, §6.1.2.2.

A4.2.4.2 Off-Resonance

The -3dB points are specified by (A4-9). Additionally, we can simplify the expression by stating that we will apply only a small amount of damping, that is, $Q_0 \Omega \gg \theta \Rightarrow Q_0 \gg \theta$. With that latter condition, the Ω^2 term can be ignored, as discussed in §A4.2.2. We can now write the expressions quoted in §6.3.1.2ff.

A4.3 Noise Factor of Series-tuned Antenna

A4.3.1 3dB Bandwidth

The 3dB signal bandwidth is found by considering

$$\left| \frac{R_d}{R_d + \eta^2 Z} \right|^2 = 1 / \left| \frac{\theta}{\theta - 1} + j \frac{Q_0 \bar{\Omega}}{\theta - 1} \right|^2 \quad (\text{A4-16})$$

and, if $\theta \gg 1$, this becomes

$$\left| \frac{1}{1 + j Q_0 \bar{\Omega} / \theta} \right|^2 \quad (\text{A4-17})$$

showing that the 3dB bandwidth has the same form as the same as (A4-9), namely

$$Q_0 \bar{\Omega} = \theta \quad (\text{A4-18})$$

Inspection of (A4-16) shows that this expression for bandwidth is true for all θ .

A4.3.2 General Expression

We reanalyse (7-4) with the inclusion of a series damping resistor R_d across the amplifier input, writing the antenna impedance as Z . We note that the amplifier will see the fraction $R_d / |R_d + \eta^2 Z|$ of the signal voltage, and the same fraction of the thermal noise and atmospheric noise. Additionally, the noise current will flow in R_d and in the apparent source impedance $\eta^2 Z$. Applying these parameters we obtain

a) Signal voltage $\omega \mu H \cdot \Phi \sqrt{R} \cdot \eta \cdot \frac{R_d}{|R_d + \eta^2 Z|} \quad (\text{A4-19a})$

b) Thermal noise $\sqrt{4kT_0 BR} \cdot \eta \cdot \frac{R_d}{|R_d + \eta^2 Z|} \quad (\text{A4-19b})$

c) Atmospheric noise $\sqrt{4kT_0 BF_a R_r} \cdot \eta \cdot \frac{R_d}{|R_d + \eta^2 Z|} \quad (\text{A4-19c})$

d) Amplifier noise voltage $\hat{v}_n \sqrt{B} \quad (\text{A4-19d})$

e) Amplifier noise current in R $\hat{i}_n \sqrt{B} \cdot \left| \frac{R_d \eta^2 Z}{R_d + \eta^2 Z} \right| \quad (\text{A4-19e})$

which results in

$$F_n = 1 + (F_k \Phi)^2 + \frac{P_n}{4kT_0} \left(\frac{R_n}{R \eta^2} \frac{|R_d + \eta^2 Z|^2}{R_d^2} + \frac{1}{R_n R \eta^2} |\eta^2 Z|^2 \right) \quad (\text{A4-20})$$

Noting that the impedance Z is $R(1 + jQ_0\bar{\Omega})$ and substituting

$$\theta = 1 + \frac{R_d}{\eta^2 R} \quad (\text{A4-21})$$

leads to an expression for the noise-matching term of

$$A_n = \frac{R_n}{R \cdot \eta^2} \left(\left(\frac{\theta}{\theta - 1} \right)^2 + \left(\frac{1}{\theta - 1} Q_0 \bar{\Omega} \right)^2 \right) + \frac{R \cdot \eta^2}{R_n} \left(1 + (Q_0 \bar{\Omega})^2 \right) \quad (\text{A4-22})$$

which, with further approximations, leads to the expression noted in the main text, §6.3.2.

A5 Sequence Design

A5.1 Fourier Transform Pairs

The discrete Fourier transform (DFT) pair of $f(nT)$ and $F(m\Omega)$ we define such that

$$\begin{aligned} f(nT) = 1 &\Leftrightarrow F(m\Omega) = N \delta \\ f(nT) = \delta &\Leftrightarrow F(m\Omega) = 1 \end{aligned} \quad (\text{A5-1})$$

or, more explicitly,

$$\left. \begin{aligned} f(nT) &= \frac{1}{N} \sum_{m=0}^{N-1} F(m\Omega) e^{jmn\Omega T} \\ F(m\Omega) &= \sum_{n=0}^{N-1} f(nT) e^{-jmn\Omega T} \end{aligned} \right\} \quad (\text{A5-2})$$

with $\Omega T = 2\pi/N$.

A5.2 Convolution Theorem, and Correlation

We will use ‘normalised’ correlation and convolution operations – e.g. the right-hand side of (9-1) is δ rather than $N \delta$. The convolution theorem takes the form of

$$\mathbf{DFT}(f \otimes g) = \frac{1}{N} \mathbf{DFT}(f) \mathbf{DFT}(g) \quad (\text{A5-3})$$

where f and g are periodic functions of N discrete samples, the convolution is applied over one period of the functions, and \mathbf{DFT} is the discrete Fourier transform defined above. Normalised convolution implies that a factor of $1/N$ is assumed in the convolution on the left hand side of the equation, hence the factor of $1/N$ on the right hand side.

We are using the operation of correlation, \oplus , rather than convolution, \otimes , so a time-reversal is required. Denoting the time reversal of a sequence by a tilde (\sim) over the sequence, we will use a definition of correlation that results in

$$\begin{aligned} a \oplus b = c &\Leftrightarrow \tilde{a} \otimes b = \tilde{c}, \\ a \otimes \tilde{b} &= c \end{aligned} \quad (\text{A5-4})$$

It is also useful to note that for a set of real coefficients f with discrete Fourier transform F the time reversal of f can be equated with the complex conjugate F^* by

$$f \leftrightarrow F \Leftrightarrow \tilde{f} \leftrightarrow F^* \quad (\text{A5-5})$$

A5.3 Parseval's Theorem

Using the above definitions, this states

$$\sum_{n=0}^{N-1} |f(nT)|^2 = \frac{1}{N} \sum_{m=0}^{N-1} |F(m\Omega)|^2 \quad (\text{A5-6})$$

A5.4 Power Spectral Density

Consider a periodic sequence in the time domain, $f(nT)$. If this is envisaged as a voltage signal driving a load of 1Ω then the mean power dissipation is simply the mean-squared signal level over N samples; that is,

$$P = \frac{1}{N} \sum_{n=0}^{N-1} |f(nT)|^2 = \frac{|f|^2}{N} \quad (\text{A5-7})$$

where we define $|f|^2$ to be shorthand for the summation. By Parseval's theorem (§A5.3), we can write this as

$$P = \frac{1}{N^2} \sum_{n=0}^{N-1} |F(m\Omega)|^2 = \frac{|F|^2}{N^2} \quad (\text{A5-8})$$

We can consider this power to be distributed over the Nyquist bandwidth of $1/2T$ and so the *power spectral density* (PSD) is $\frac{1}{2} P / T$ where T is the sample period.

Now consider the discrete Fourier transform of the auto-correlation of f . Applying the convolution theorem, and noting the relationship between time-reversal and complex conjugation (§A5.2) we can obtain

$$\mathbf{DFT}(f \oplus f) = \frac{F \cdot \tilde{F}}{N} = \frac{F \cdot F^*}{N} = \frac{|F|^2}{N} \quad (\text{A5-9})$$

and so we have a Fourier Transform pair of

$$f \oplus f \leftrightarrow |F|^2 / N \quad (\text{A5-10})$$

which we can write in terms of the PSD function, using (A5-8), as

$$\mathbf{ACF}(f) \leftrightarrow 2NT \cdot \mathbf{PSD}(f) \quad (\text{A5-11})$$

showing that we have confirmed the Weiner-Khintchine theorem, which states that the auto-correlation function (ACF) and the power spectral density (PSD) function form a Fourier transform pair. In our particular case, there is a factor of $2NT$ to consider.

The systems we are considering are of the form $f \oplus h \rightarrow r$ or, in the frequency domain,

$$\left. \begin{aligned} \frac{F(m\Omega) \cdot \tilde{H}(m\Omega)}{N} &\rightarrow R(m\Omega) \\ \frac{\tilde{F}(m\Omega) \cdot H(m\Omega)}{N} &\rightarrow \tilde{R}(m\Omega) \end{aligned} \right\} \quad (\text{A5-12})$$

which we can multiply together, to give

$$|F|^2 \cdot \frac{|H|^2}{N^2} \rightarrow |R|^2 \quad (\text{A5-13})$$

The function $|F|^2$ is proportional to the *power spectral density* (PSD) function, discussed above, and so the system transfer function, for power spectral density, is $|H|^2 / N^2$. If the input is represented by a PSD function $G(m\Omega)$, then the output PSD is

$$G(m\Omega) \frac{|H(m\Omega)|^2}{N^2} \quad (\text{A5-14})$$

This is a familiar result, but modified by the inclusion of the N^2 term due to the fact that we are considering a periodic sampled system with particular definitions of Fourier transform and normalised convolution.

A6 Cave Locations for Trials

The channel sounder is considered to be a fully-portable piece of equipment but, in the early stages of this project, we envisaged having to choose a cave where there was a convenient source of mains power, for the computer equipment that would be used to log the results. Several locations were identified in the Yorkshire Dales, as follows.

White Scar Cave: This is a show cave just outside Ingleton on the B6255 to Hawes. The cave entrance is at 260m o.d., and rises gently over its 2km length. A central section runs under a limestone bench spur to the WSW of Ingleborough. The far end of the cave is close to the surface, in the vicinity of the Fell Lane track to Ingleborough. Inspection of the OS map and a survey of the cave suggest that a section of the cave between Swift's Lake and the Hall of Justice – around 1 to 1.5km into the cave – lies below the limestone pavement at 420m - 430m o.d., and is therefore at a depth below the surface of up to 160m. Traversing the pavement SE or NW will reduce the transmission depth whilst increasing the offset. Traversing NE towards Ingleborough will maintain the depth whilst increasing the offset of the field point.

Stump Cross Caverns: This is a show cave on the B6265 between Grassington and Pateley Bridge. The cave, although extensive, does not go very deep, and the surface above is fairly flat. It is unlikely that this will prove to be a worthwhile location for experiments.

Bull Pot Farm: This is a self-catering bunk-house owned by the Red Rose Cave & Pothole Club. It is situated in a remote area, at the end of the minor road from Kirkby Lonsdale to Casterton Fell, but it does have electricity and telephone. It is close to the Lancaster Hole / Easegill Caverns / Pippikin Pot cave system which is one of the largest in the country. There are many entrances within 2km, but the caves are not particularly deep. This location would be particularly suitable for experiments with radio-location because of the extensive area of the cave system and the accurate surveying that has already been done in this system.

Of these locations, White Scar Cave was thought to be the most suitable for the initial channel-sounding experiments because

- i) With around 1km of cable we could place the receiver in a 'deep' location, at 160m below the surface.
- ii). The surface above the receiver is common land with no restrictions on access. It is limestone pavement so there would be few problems due to the screening effect of a wet peat overburden.

- iii) The cave is privately owned with access not normally granted to cavers. This was seen an advantage because the equipment and cabling would not be disturbed. The cave has previously been made available for cave radio experiments.

In addition to these locations, we have prepared a list of 'deep' caves in the Yorkshire Dales, (Gibson, 1994). Our definition of deep is different to that used by cavers (who tend to measure depth vertically from the entrance) and the media (which tends to measure it horizontally²). Instead, we mean 'deep' to mean 'surrounded by this distance of rock in all directions'. This is a more difficult quantity to measure, because caves whose entrances are high in the hills, often resurge in the valleys, so the lower reaches of a cave can be quite shallow.

A6.1 References

Gibson, D. (1994). Sites for Testing Cave Radios at Depth. *CREGJ* 18, 12-14

2 As in "cavers trapped two miles underground!"

A7 Channel Sounder Design

A7.1 Design Status

From the notes below, it can be seen that this project is clearly still only in a very early state of development.

A7.1.1 Hardware

The design is still in a development phase, and has not been completed. With reference to the block diagram, *Figure A7-2*, on page 252, the receiver pre-amp board has not been constructed, and initial tests have been done using a low-gain pre-amp on the DSP board. The transmitter has not been constructed, meaning that only measurements of noise and interference are currently possible. And, because the transmitter has yet to be constructed, the receiver is not synchronised using the proposed code-locked loop. Instead, the receiver operates from a 16MHz clock on the DSP board.

A7.1.2 Software

Simple routines have been written to exercise the receiver DSP board, to allow it to collect data and log it to the hard disc. However, no attempt has been made to write a user interface, utilising the keypad and LCD, and no event-driven data-logging routines have been written. The TDS microcontroller is currently controlled from a PC with a serial data link – i.e. no ‘turnkey’ routines have been written.

A7.2 TDS 2020 Motherboard

This Eurocard-sized board is the interface between the TDS 2020 microcontroller card and the rest of the system. We will give just a brief description of its operation. A circuit diagram is shown in *Figure A7-3* on page 254 and a PCB assembly diagram in *Figure A7-4* on page 255.

The board takes 12V d.c. power through a reverse-protection mosfet (Q3) and supplies it to the TDS card, where it is regulated and then brought back to this motherboard for supplying the other boards in the set. Components around IC4-A provide a temperature-compensated bias for the LCD module³. Q1/X1 provide an on-board sounder in case any beeps are needed. PL2 is the programming interface to the TDS card. IC5 and 6 are the matrix keypad driver⁴. Components around IC4B include two temperature sensors R5 and Q2. This is simply to give the board

3 This experimental circuit did not function as expected.

4 This keypad driver does not function with built-in TDS 2020F commands. It is believed that this is due to a slow rise time on the signal, caused by the high value of RP1, but this had not been verified.

some functionality during initial testing. SK6 is the connection to the LCD module. PL4 is the connection to an external keypad. Some key switches are included on the board. PL3 is the ribbon cable connection to the rest of the system.

The board includes some components, based around IC1 and IC2, which are not normally fitted. These are only used if an earlier non-flash ROM version of the TDS 2020F is used, when they replace the TDS's programming daughter-board. (It is not possible to use TDS's supplied daughter board when the hard disc PCMCIA card is fitted). When external memory is fitted the link LK1 must be moved to the 'EXT' position.

Apart from the two problems mentioned in footnotes 2 and 3 above, we have experienced some problems with the current consumption of this board. The TDS 2020F is intended to shut down to a standby current of under $200\mu\text{A}$, but the quiescent current is several mA. Experiments have shown that it is caused by one of the components on this motherboard, but we have so far failed to identify the exact cause. We have not yet built a second prototype but, during its eventual construction, we will investigate this problem in more detail.

A7.2.1 External Interface

To use the channel sounder, the following connections are made to the TDS 2020F motherboard.

- i) 12V power
- ii) RS232 serial data link
- iii) TDS 2020F card (with hard disc card plugged into TDS card)
- iv) ribbon cable connection to receiver DSP board
- v) Keypad (not yet functional)

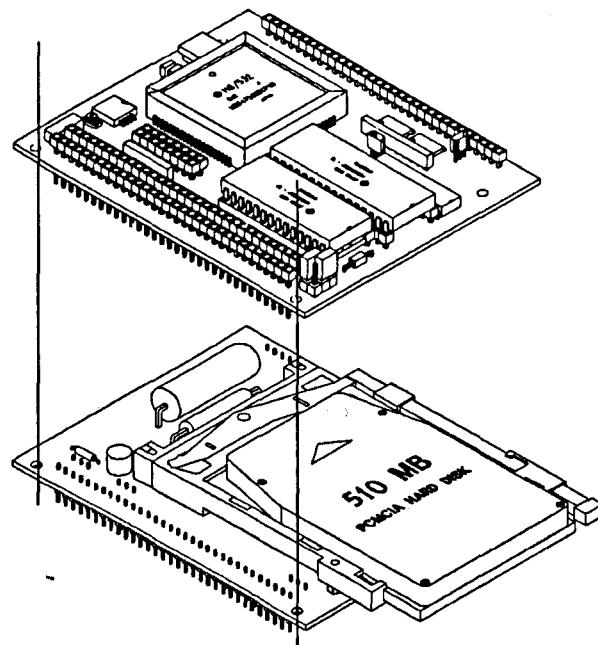


Figure A7-1 : TDS Microcontroller and 500Mbyte hard disc card.

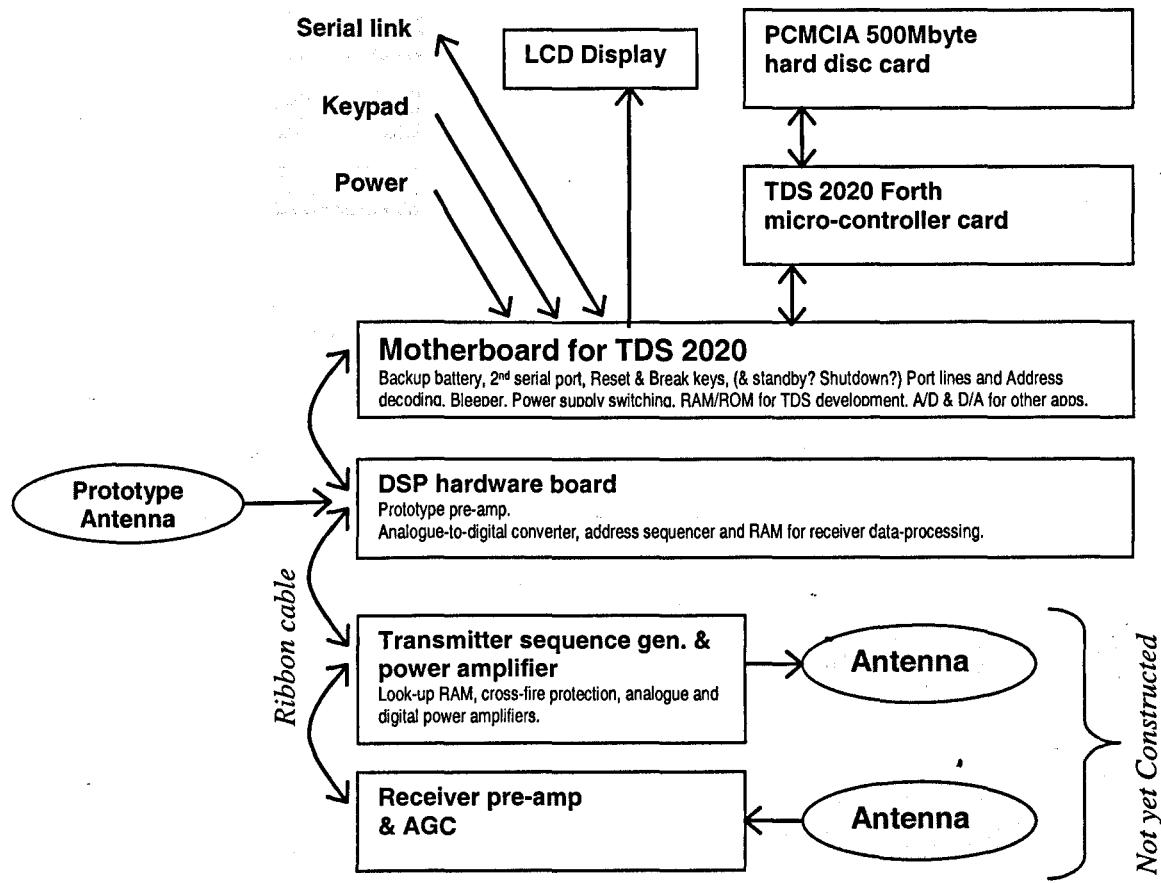


Figure A7-2 : Interconnection of channel sounder modules.

A7.3 DSP Capture Board

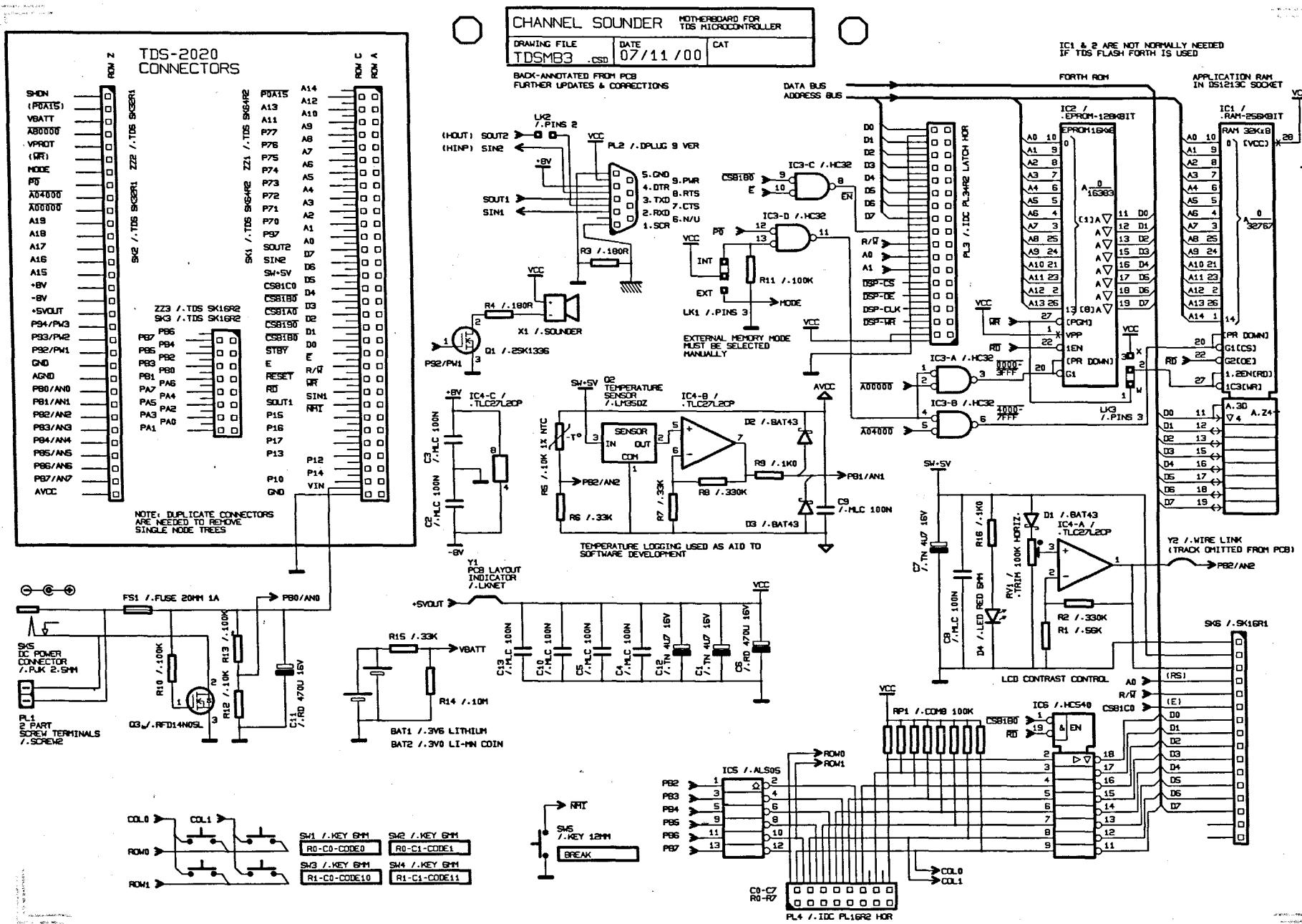
The DSP board performs data capture and signal averaging in hardware. At the time this project commenced, this was considered to be the best approach, but it lead to unforeseen difficulties and, if the project were to re-commence, we would adopt a software solution using a dedicated DSP microprocessor. This module captures data at high-speed from an ADC, and stores it to RAM. It is currently intended to be used with an AD7721, which produces 16-bit samples at a rate of 500kHz. A single data set comprises 2^{17} (128K) readings (i.e. 256K bytes). The module is designed to capture 128 data sets in a single operation and to perform signal averaging across the data sets in real-time. The result of the signal averaging is that 128K words of 4 bytes are stored in the RAM and then transferred to a 500Mbyte hard disc card. In summary, the data capture and processing routines are as follows

- 16 bit ADC is clocked at 16MHz. Data is sampled at 500kHz / $2\mu\text{s}$
- FIR filter, built-in to AD7721, results in -3dB bandwidth of 245kHz.
- Data is handled as 32-bit (four byte) values, and is stored in a 4Mbit (512Kbyte) RAM.
- 128K samples (at 500kHz) are collected in 262ms.
- 128 data sets are hardware-averaged in 33.6s
- Data is written to 500Mb hard disc card

The intention is that the DSP card is programmed by the TDS microcontroller to act as a data logger. It is envisaged that the equipment can be left unattended underground for some time, and the data card eventually retrieved for processing on a PC.

The digital section of the board is shown in the form of a block diagram in *Figure A7-5* on page 256. The heavy line in the figure shows the main signal path during the 'RAM-read' part of the cycle, when data is read from the RAM and added to the ADC sample. The next phase of the cycle is where the data is written to the RAM. This 'read-accumulate-write' cycle forms the basis of the real-time signal averaging function. The RAM is addressed by a 19 bit bus, and uses 4 bytes to hold each data sample. It therefore holds 2^{17} samples, or 131,072. A 7-bit counter can be programmed to count the number of datasets that are signal-averaged, from 1 to 128.

Figure A7-3 : Circuit diagram of microcontroller motherboard.



tdsmb-as.cdr 4/10/00

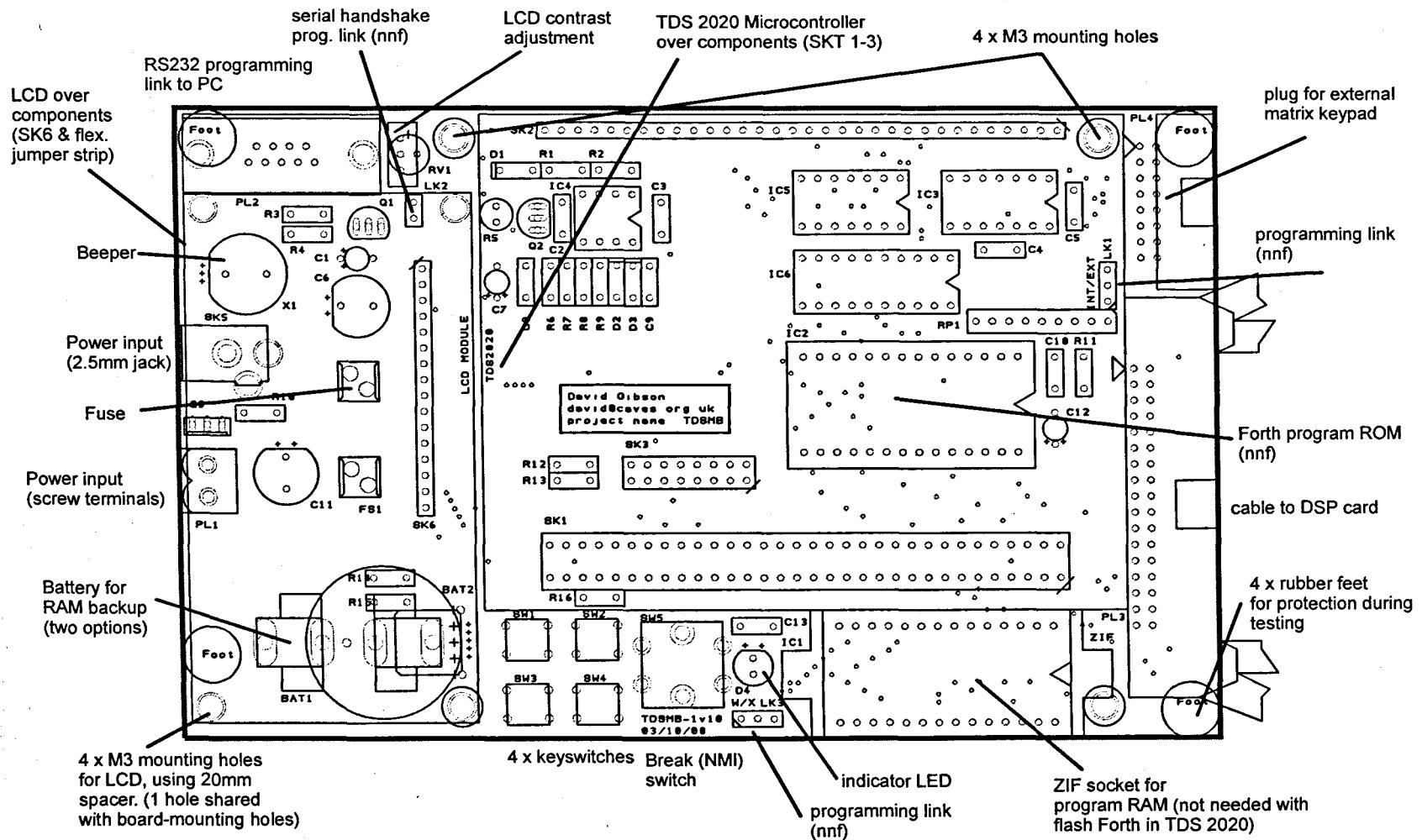


Figure A7.4 : Assembly diagram of microcontroller motherboard.

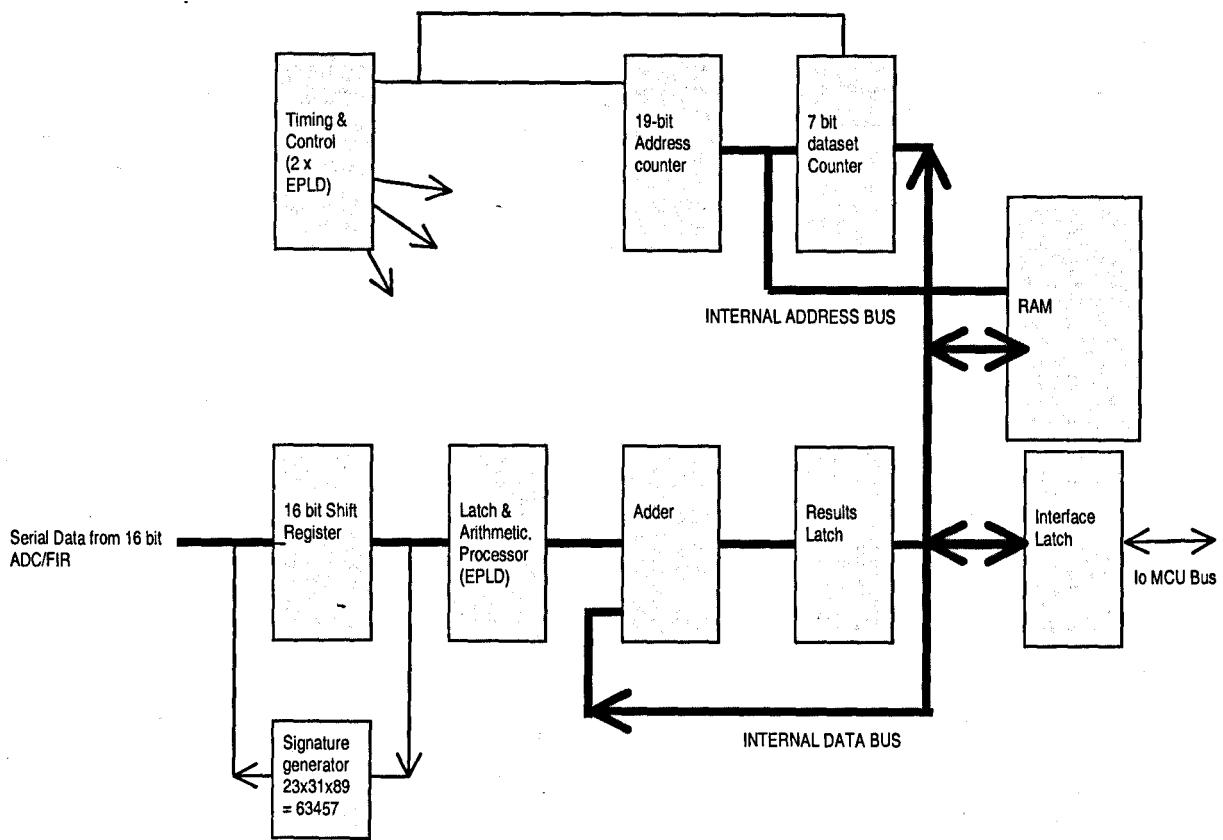


Figure A7-5 : Channel sounder DSP board; block diagram.

A7.3.1 Description of Digital Section

One important feature to note is that the circuit diagrams make extensive use of IEEE logic symbols (Kampel, 1986), which allow the EPLD operations to be described succinctly on the diagram. The circuit diagram is shown in **Figure A7-6** on page 260.

The salient point of this module is a real-time hardware signal-averaging routine. The input to this module is the amplified and filtered analogue signal from the receiver pre-amplifier. This is digitised, with 16 bit samples being produced at a rate of 500kHz. As each sample is received, the corresponding cumulative sample is retrieved from the RAM, added to the current sample and stored back in the RAM. This ‘read-sum-write’ cycle operates at 500kHz. The RAM addresses are generated by a hardware counter, and a second counter counts the number of datasets that have been summed. The operation is controlled by a logic sequencer designed around three small EPLDs (eraseable, programmable logic devices).

A7.3.1.1 Performing the signal averaging

The 16-bit data is clocked, serially, at 16MHz into a shift register, IC17-18, and then multiplexed onto an 8 bit bus. With no further processing, each data sample would require 2

bytes of storage in the RAM, IC6. However, we want to average 128 data-sets, each containing 131,072 samples. To avoid the need for a temporary storage of 32Mb the signal averaging is performed in real time, by reading a value from the RAM, adding the current data and writing the result back to the RAM. In this way, only 2^{17} (128K) values need to be stored but, because 128 data-sets are being added together, each value will require three bytes. In fact, each cumulative addition is performed as a four-byte operation, for ease of design, so 128K samples takes up 512Kbytes of RAM, in a single 4Mbit chip.

The addition is performed by the adder chips, IC3 and 4. A byte of data is read from the RAM, added to the data from the ADC shift registers, and the result is latched into a data buffer IC5 before being written back to the RAM. The addition spans four bytes, with each ‘read-sum-write’ operation taking 250ns. The EPLD, IC2, controls the multiplexing, sign-extend and carry in/out operations. It takes 1 μ s for the 16 bits of data to be shifted out of the ADC, so the data processing takes 2 μ s in total – and the ADC provides samples at this interval.

If an ADC with a parallel output were used, then the throughput could be doubled to allow for samples at 1MHz. If only 8 bit samples were to be processed then 2MHz sampling would be achievable. These changes can be implemented, in the main, by changes to the three programmable EPLD devices on the board.

IC8 to 11 form a 16 bit counter that holds the sample number (0-65536), and which addresses the RAM. IC12-13 form an 8 bit counter that holds the msb of the sample number, and the data-set number (0-128). Timing and control is provided by EPLDs IC7 and IC16

The 16MHz clock produces (via IC7 etc) a timing period of 250ns. Four of these are ‘idle’ periods whilst 16 bits of data are clocked out of the ADC at 16MHz. The data addition uses the remaining four time slots in the 2 μ s ADC sample time. Although each time slot is 250ns in length, it has to comprise a read, add and write portion. Even using fast memory and AC-type CMOS the propagation delays are still a problem – especially the ‘carry’ signal that ripples through the counter chain. For this reason, IC16-C and IC16-D are novel ‘ripple traps’ which limit the distance the ripple-carry has to propagate, and so speed the operation of the counter. A side-effect of this is that the counter chain does not increment in a normal binary sequence but, since all it does is address the RAM, this is not a serious problem.

A7.3.1.2 *Reading the data*

Once the data is stored in the RAM, it can be clocked out, under the command of a micro-controller, and sent for further processing. The obvious way to do this is by placing the RAM on a microprocessor bus, but this would require extra tri-state buffers; and the non-sequential nature of the address counter (mentioned above) becomes a problem. Instead, the RAM is

cycled through a ‘dummy’ data capture routine, and the RAM data is read through latch IC15 onto the microprocessor bus. The details of this function is taken care of in the EPLDs.

A7.3.1.3 Testing

To check that the board is operating properly, we could capture a pre-determined sequence of data and check that the data eventually read from the RAM agreed with a template stored in the MCU. IC19-A to IC19-C are used as part of a signature-generation circuit. Asserting TEST replaces the ADC data with a sequence generated using the ex-or gates. Unfortunately, the shift register outputs are bus-multiplexed and not all accessible for a simple PRBS generator. However, bits 0, 7, 8 and 15 are available continuously and, with the feedback given by $\neg(B0 \oplus B7 \oplus B8 \oplus B15)$ and the shift register clearing to zero, the sequence has 63457 states. The resultant fairly random RAM contents can be read into the MCU, which can then perform a signature analysis, although strictly speaking, this is not a true signature analysis, because the sequence is not pseudo-random.

A7.3.1.4 Control Signals

The operation of the module is controlled by a ‘register’ of signal lines. Both the register and the data i/o are memory-mapped to the MCU. The control lines are $\sim EN$, $R/\sim W$; there are two address lines $AX0:1$ and a range of other signals.

$DX0-DX7$	8 bit i/o data bus, connected direct to MCU bus on micro-controller card.
$\sim EN$	Enable. Decoded with $R/\sim W$ to address registers. $EN = CS8180 \& E$, from the MCU’s control lines.
$R/\sim W$	The MCU’s read/write line
$AX1:AX0$	The MCU’s $A1:A0$ lines
$\sim CSTX, \sim OE, CLK, \sim WRTX$	Control signals generated on this board and sent, via the edge connector, to the Transmitter board. See later for a description.

In addition, there are 20 internal address lines, used to access the RAM, which are available on a through-board connector for use by the piggybacked Transmitter board. They are...

$A0$	normally a 2MHz clock (i.e. 250ns HI, 250ns LO)
$A1$	1MHz
$A2A$	500kHz
$A2:A18$	250kHz:3.8Hz

The control lines are used to access two registers for reading, and two for writing, namely

AX1:AX0	R/$\sim W$	Function
00	0	WRX0 active, write to dataset register
01	0	WRX1 active, write to control register
00	1	RDX0 active, read from data register
01	1	RDX1 active, read from status register

Other values of AX1:AX0 are used to address other boards in the assembly.

The data register contains...

GO	This line is taken high to initiate data collection. It must return low before one dataset has been completed.
TEST	This causes the ADC data stream to replaced by a test sequence (a 16 bit non-maximal length sequence).
READ	This is low for data capture. If READ is taken high before GO is pulsed then a 'dummy' data-set capture occurs, using an external clock, STEP instead of the internal 16MHz clock. During these dummy cycles the RAM data can be read on DX0-7 when ~RDXL or ~RDXH are low.
WRITE	In conjunction with READ this allows us to scan the data receiver, or the look up table for the data transmitter. Its effect is to alter the way the internal address counter operates.
STEP	When READ or WRITE is high, pulsing this line will increment the address counter.
CLKOFF	Disables on-board crystal oscillator
CAL, STBY	control lines for the ADC.

The status register contains...

BUSY	this goes high in response to GO, and stays high until all the data sets have been collected.
DVAL	status lines for the ADC. Details to be written.

The dataset register is programmed with the two's complement of the number of data-sets to be averaged. That is \$00 represents 128 datasets, and \$FE represents 1 data set. (The lsb is not used and must be written as zero).

A7.3.2 Description of Analog Section

The circuit diagram is in *Figure A7-7* and *Figure A7-8* on pages 261 and 262.

A7.3.2.1 Signal Conditioning

This data-capture board is intended to be used with a separate analogue signal-conditioner. However, for early trials, a trimmed-down signal conditioning 'front-end' is available on the data capture board. This is based on an industry-standard three op-amp instrumentation amplifier, but there are some points of interest, which we should note. A standard instrumentation configuration would utilise a third, summing, amplifier after IC23-A and IC24-A. This summing amplifier is omitted because the ADC has differential inputs. Converting the signal from differential to single-ended would compromise the CMRR. The amplifiers operate from a single supply rail, and must provide differential outputs with the correct bias to drive the ADC. IC25-A provides this bias which is derived from the same reference voltage that is used by the ADC. *Figure A7-10a* shows a conventional instrumentation amplifier. Note that the configuration used on the data-capture board is modified by the addition of extra feedback components, as shown in . For the standard configuration, the differential gain is easily calculated to be $1 + (R1 + R3)/R2$. For the enhanced circuit the analysis is slightly more complicated and results in a high-pass filter action. We can see that if R1 and R3 were omitted,

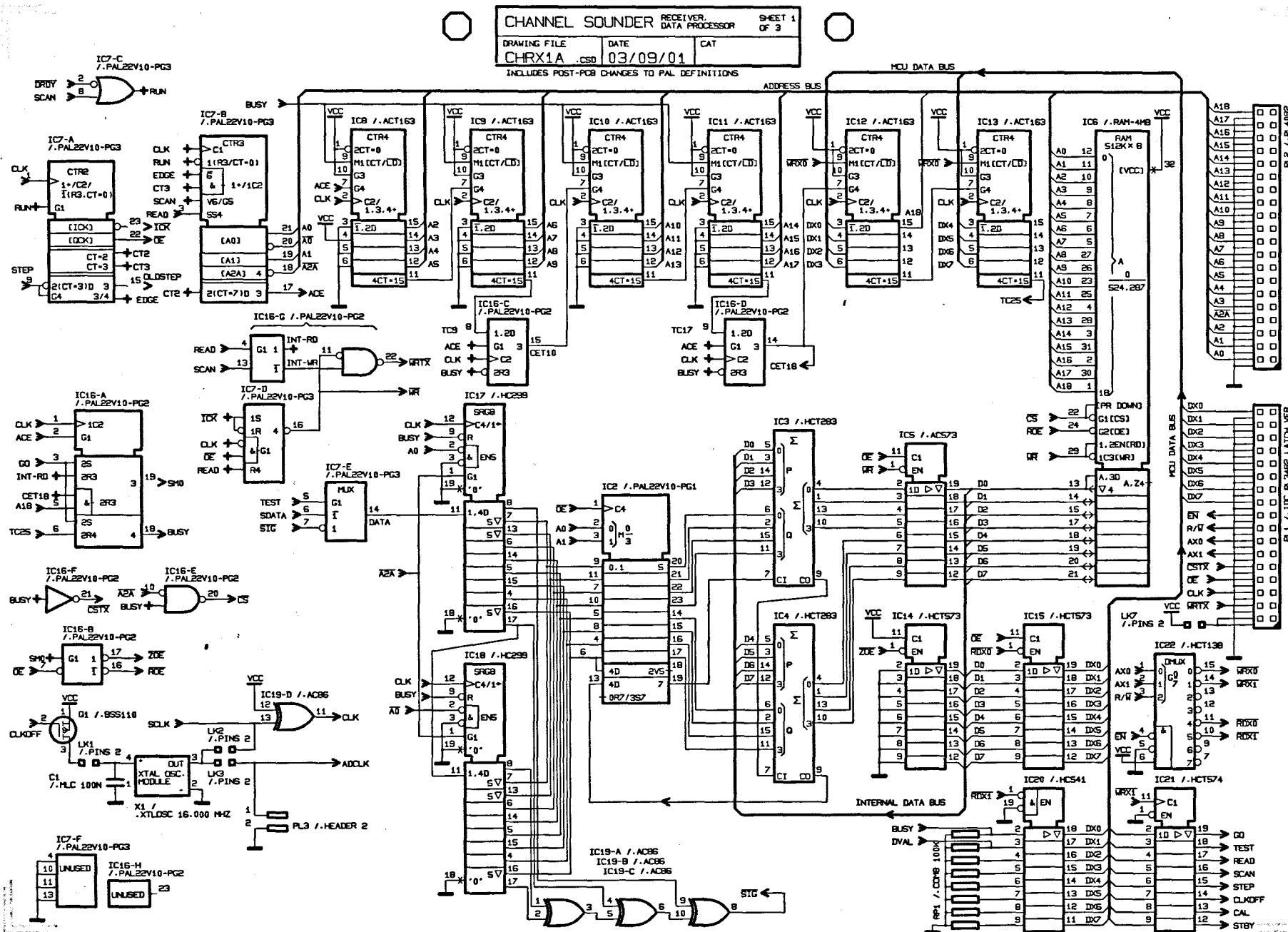
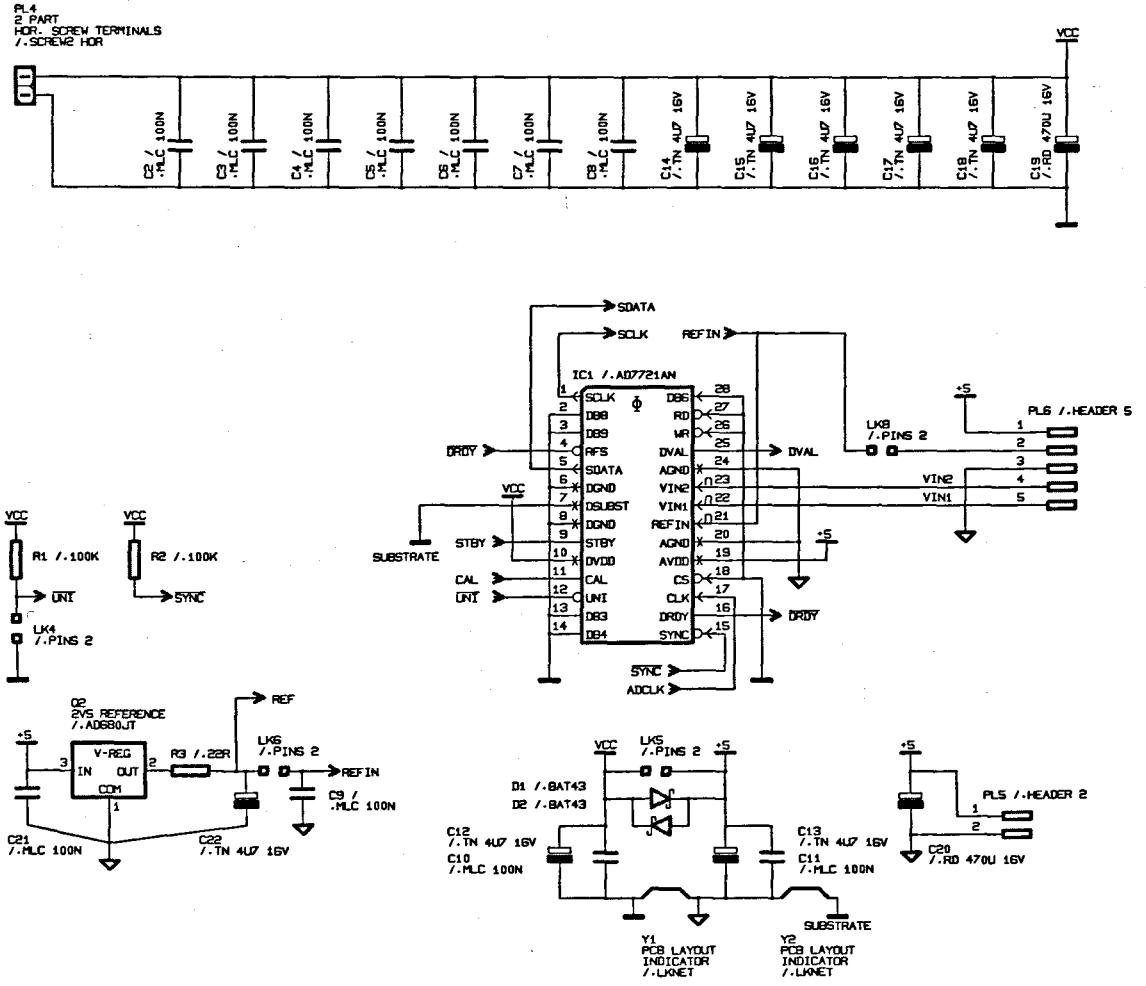
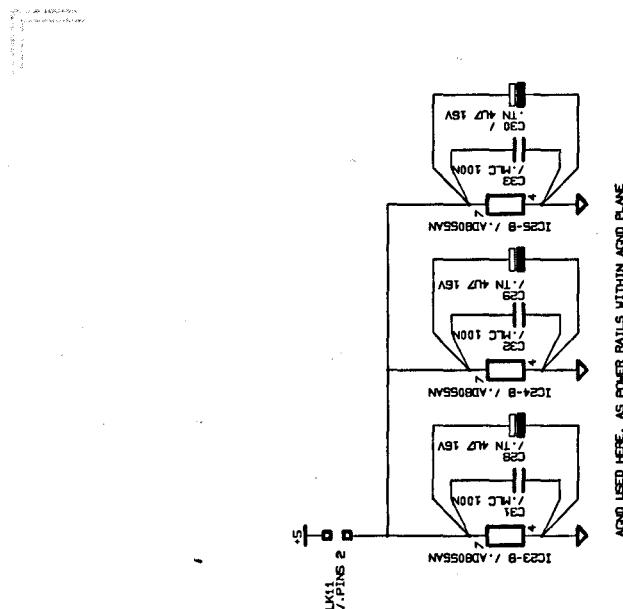


Figure A7-6: Circuit diagram of DSP/capture board, sheet 1 of 3.
This shows the digital section of the DSP board.

CHANNEL SOUNDER		RECEIVER, DATA PROCESSOR	SHEET 2 OF 3
DRAWING FILE	DATE	CAT	
CHRX1B .CSD	14/02/01		

SHEET B - ANALOGUE & POWER SUPPLY CIRCUITRY





SHEET C - TRIAL RECEIVER FRONT-END

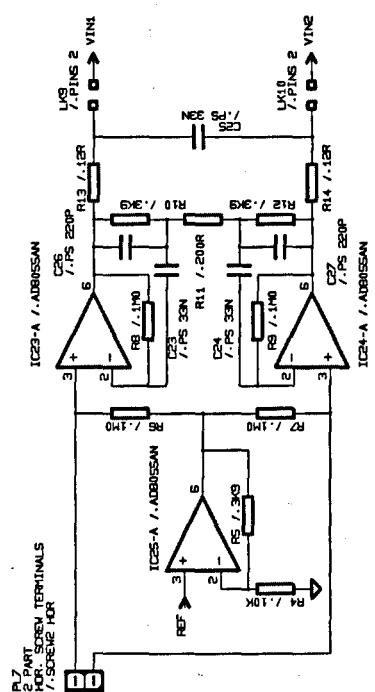


Figure A7-8 : Circuit diagram of DSP/capture board, sheet 3 of 3.
 This shows the simple fixed-gain pre-amp used during preliminary tests.

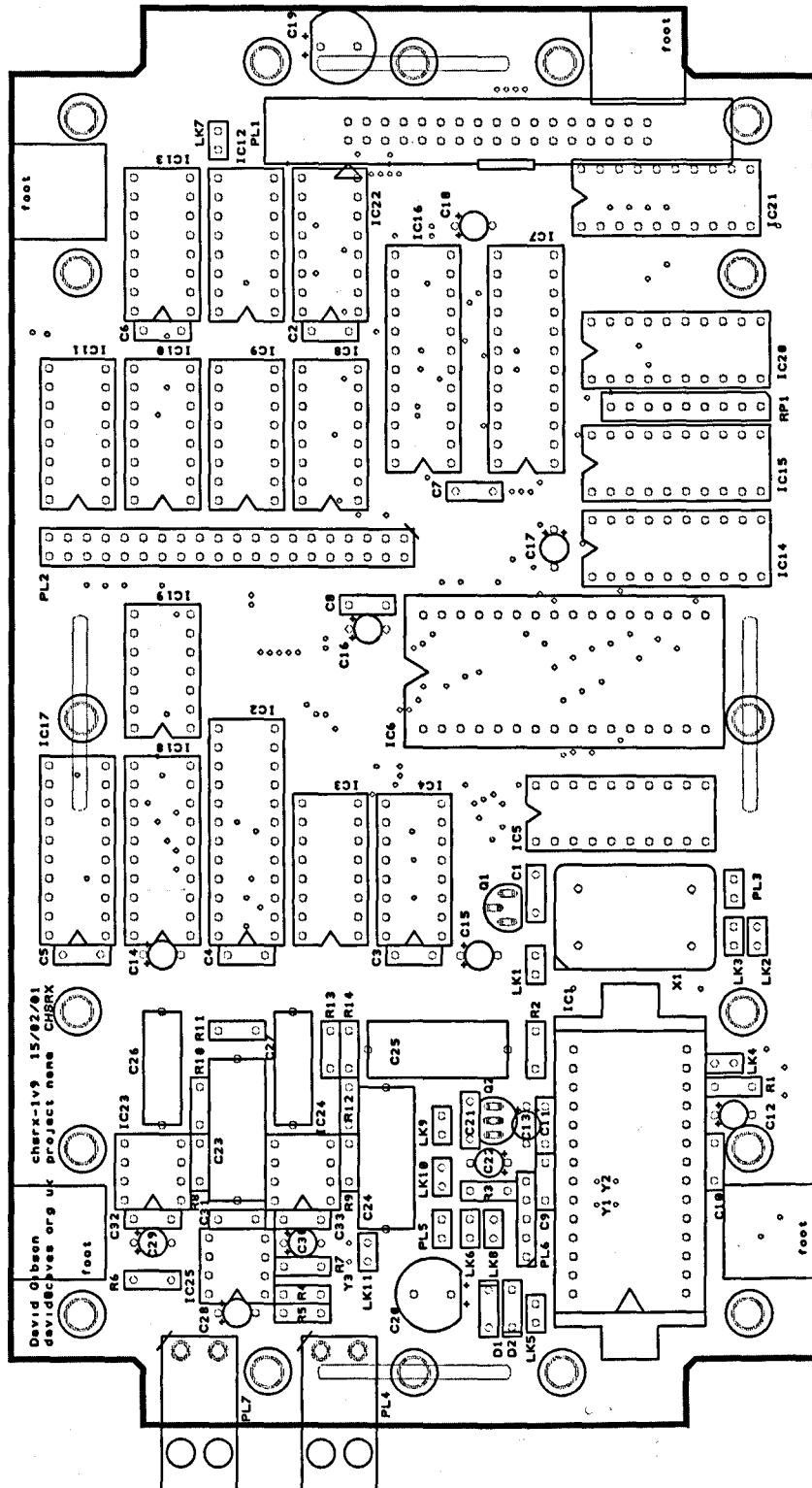
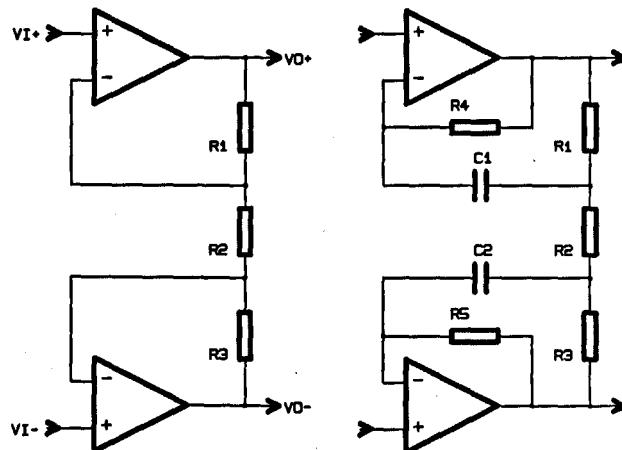


Figure A7-9 : Assembly diagram (un-annotated) of DSP/capture board..

the gain would simply be $1 + (R4 + R5)/R2$ with a high-pass time constant of $(C1 + C2)R2$. However, for the thermal noise of the circuit to be low, $R2$ must be low which requires a large value for the capacitors. Given that we might consider polystyrene capacitors for lowest distortion, this would result in a physically large and expensive component. However, what the analysis shows is that with $R1$ and $R3$ present, provided $R4 \gg R1$, $R1 \gg R2$ etc, the circuit behaves as if the value of $C1$ were multiplied by the ratio $R3/R1$, and similarly for $C2$. With the component values given, we therefore have a differential gain of 40 and a low frequency cut-off of 190Hz, caused by $C1$ appearing to be 250 times as large as its true value.

The 200pF capacitors $C26$ and $C27$ form a low-pass filter at 185kHz, and a second low-pass filter is formed by $R13$, $R14$ and $C25$, at 200kHz. These form the anti-aliasing for the ADC.

The op-amp chosen – the AD8055 – is a low-cost high performance device. It has a low noise of $6nV/\sqrt{Hz}$ and $2pA/\sqrt{Hz}$ which is flat above 10kHz. This is not the lowest noise that can be achieved. The open-loop gain is 4000 at 1MHz, with a gain-bandwidth product of 300MHz. At 100kHz the quoted figure for distortion is $-100dBc$. This performance is typical of the latest generation of video-bandwidth devices and, being a voltage feedback amplifier it is easy to use.



*Figure A7-10 : Instrumentation amplifier.
(a) standard configuration, (b) the author's design, with capacitance multiplier.*

A7.3.2.2 Analogue to Digital Conversion

The analogue-to-digital conversion is performed by IC1, an AD7721. This is an advanced device, with built-in over-sampling, decimation and a 211-tap FIR anti-aliasing filter. It produces 16-bit samples at 500kHz from a 16MHz over-sampling clock, with a $-3dB$ bandwidth of 244kHz which is very close to the Nyquist limit. Its response at 276kHz is $-72dB$. The FIR filter considerably eases the anti-aliasing requirements. With signal-conditioning first-order low-pass filters at 185 and 200kHz we can say, to a good approximation, that the attenuation at 16MHz will be 1/80 or $-38dB$. In order to sample at 500kHz, the AD7721 must be clocked at 16MHz, which is slightly outside its spec. of 15MHz.

A7.4 EPLD Programming

The logic equations for the EPLDs used on the channel sounder board were compiled using one of two programs.

- i) **PLPL**, an MS-DOS program, PLPL 2.1, written by the Applications Department at Advanced Micro Devices (AMD) in 1986. Such a program has a very low functionality by today's standards but is still suitable for the basic task of compiling simple logic equations and generating simple test vectors.
- ii) **Synario**, a suite of software from Vantis (www.vantis.com) intended for programming their MACH devices, it includes a generic compiler.

A7.4.1 PG1 – data latch

The logic functions for this device are straightforward, corresponding to the IEEE logic symbol for IC2 in the circuit diagram in *Figure A7-6* on page 260. A program listing is included on the attached CD-ROM. (See appendix A9.2).

A7.4.2 PG2 – gates and control latches

The logic functions for this device correspond to the IEEE logic symbols for IC16 in the circuit diagram in *Figure A7-6* on page 260. There are several separate elements to this device, mostly concerned with simple gating or latching functions. A program listing is included on the attached CD-ROM. (See appendix A9.2).

A7.4.3 PG3 – timing generation

The logic functions for this device correspond to the IEEE logic symbols for IC7 in the circuit diagram in *Figure A7-6* on page 260. There are several separate elements to this device, which are concerned mainly with the timing generation. This device had to be redesigned several times during the development phase, thus amply demonstrating the use of re-programmable hardware. Its function is now, unfortunately, somewhat obscure and it would be advantageous, at the current stage of the development, to re-assess the control signals and to consider a redesign of the PCB. The IEEE logic symbols for IC7, shown in *Figure A7-6* on page 260 do not correspond to the logic equations listed for this device, as the diagram shows an earlier version. A program listing is included on the attached CD-ROM. (See appendix A9.2).

A7.5 References

Kampel, I. (1986). *A Practical Introduction to the New Logic Symbols*. London: Butterworths. (2nd Edition).

A8 Data Processing

A8.1 Window Functions

When calculating the Fourier transform of a finite sequence it is implicit that the sequence can be extended to infinity in a periodic fashion. But this process can lead to discontinuities at the bounds of the sequence, which have the effect of introducing additional spectral components into the signal. This phenomenon is often referred to as *spectral leakage*. It can be reduced by decreasing the sampling period, although the leakage is not caused by the sampling process itself.

Simplistically, one might assume that each coefficient or ‘point’ in the frequency spectrum represented the amount of that frequency component that was present in the original signal, but this is not so. Instead, each coefficient can be thought of as being the output from a narrow-band filter centred on that frequency. The effect of spectral leakage is to add significant sidelobes to that filter. *Figure A8-1a*, shows this effect. Weighting functions can be applied to the data to reduce the spectral leakage; these are known as ‘window functions’. The effect of one such window is shown in *Figure A8-1b*.

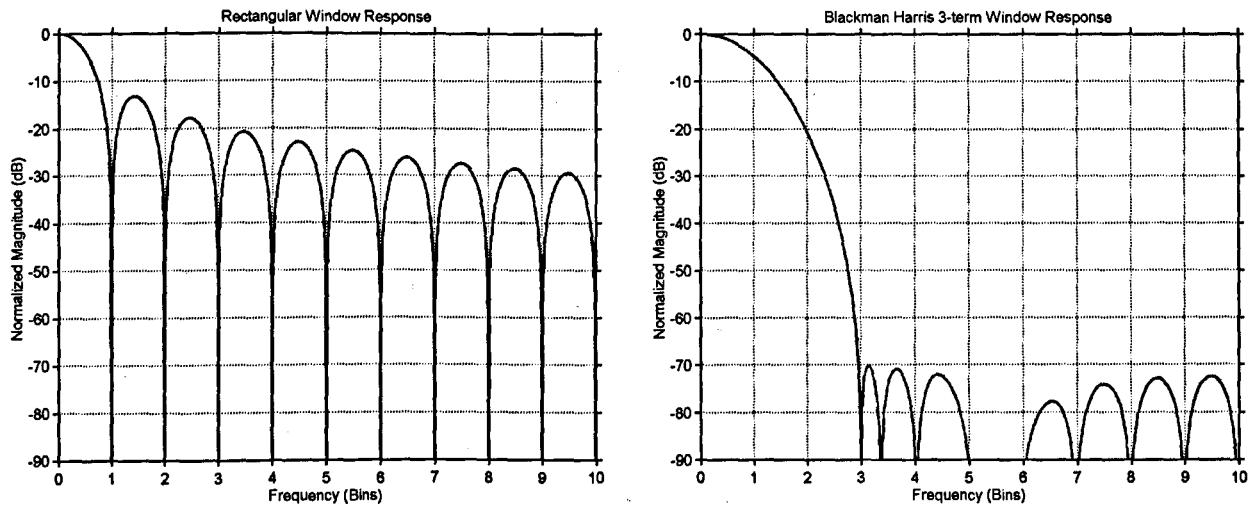


Figure A8-1 : DFT spectral response and the effect of a window function.

The DFT acts like a bank of filters, each with the comb-like structure shown. The high-sidelobes lead to spectral leakage. **a) Left:** without a window function, **b) Right:** the response after application of a minimum three term Blackman-Harris window.

A8.1.1 Generalised Cosine Windows

There are a number of parameters we can use to characterise different windows, and these are shown in *Table A8-1*, for a range of *generalised cosine windows*. This family of windows has the advantage that it is easy to calculate. We can characterise a window function by means of a number of parameters, which are summarised below and referred to in the table.

Parameter	Rectangular Window	Hamming Window	Blackman Window	Blackman-Harris 3 term	Blackman-Harris 4 term	BH4-modified Window
3dB bandwidth	0.89	1.31	1.66	1.64	1.91	1.90
6dB bandwidth	1.22	1.83	2.31	2.28	2.67	2.67
First Zero	1.00	2.00	3.00	3.00	3.98	3.99
Sidelobe height (dB)	-13.3	-42.7	-58.1	-70.2	-92.0	-89.7
NEB	1.00	1.36	1.73	1.71	2.01	2.01
CG	1.00	0.54	0.42	0.42	0.36	0.36
SL (dB)	-3.92	-1.75	-1.09	-1.12	-0.82	-0.82
FOM	1.12	1.04	1.04	1.04	1.05	1.05
Roll-off	6dB/oct.	6dB/oct.	18dB/oct.	6dB/oct.	6dB/oct.	18dB/oct.

Table A8-1 : Parameters for a range of generalised cosine windows.

The parameters are described in the text, in §A8.1.1

The data in this table, and in Figure A8-1, was obtained from a Matlab calculation using a window of length 512, and a DFT of 65536 points, so that the window could be examined and plotted at a resolution of 128 points per sidelobe.

A8.1.1.1 Sidelobe height

Clearly the higher the sidelobes, the more spectral leakage there is. The heights of successive sidelobes drops off at 6dB/octave but, if the coefficients sum to zero, the roll-off is 18dB/octave. (Harris, 1978).

A8.1.1.2 3dB and 6dB bandwidth, first zero

The 3dB bandwidth (given here in terms of ‘points’ of the transform) gives us an idea of performance but it is the 6dB bandwidth that governs the resolution of the filter. This is unlike a conventional situation, where it is the 3dB bandwidth that indicates resolution, and is because of the coherent addition inherent in the transform. (Harris, 1978). The position of the first zero in the transform is included for interest.

A8.1.1.3 Noise-Equivalent Bandwidth (NEB)

3dB bandwidth tells us how much ‘signal’ gets through a filter. NEB tells us how much wideband noise gets through. The NEB is the bandwidth of the brick-wall filter that would admit the same noise power as the filter in question. NEB is usually greater than 3dB bandwidth. Here, we express the NEB in terms of the spectral line width of the filter. Expressed numerically in this way, the reciprocal of NEB is the noise factor of the transform (i.e. output SNR divided by input SNR), which is also called the processing gain. For the windowing time sequence $w(nT)$ we define the NEB as

$$NEB = N \frac{\sum_{n=0}^{N-1} w^2(nT)}{\left[\sum_{n=0}^{N-1} w(nT) \right]^2} \quad (\text{A8-1})$$

(Harris, 1978) deduces that the figure of merit (notated *FOM* in *Table A8-1*) given by

$$1 - \text{noise-equivalent bandwidth} / 3\text{dB signal bandwidth} \quad (\text{A8-2})$$

“appears to be a sensitive indicator of overall filter performance”. Harris does not elaborate, but we can see why this f.o.m. is useful, because it is similar to the ratio of sidelobe energy to the energy within the 3dB bandwidth, and this would indicate the degree to which out-of-band noise distorted the in-band measurements.

A8.1.1.4 Coherent Gain (CG)

It can be shown that the measurement of a signal is proportional to the sum of the window terms. We define coherent gain as

$$CG = \frac{\sum w(nT)}{N} \quad (\text{A8-3})$$

The windowed plots presented in chapter 10 are all scaled by the appropriate coherent gain.

A8.1.1.5 Scalloping Loss (SL)

This figure represents the loss of signal due to the comb-like structure of the windowed frequency response. If a signal frequency lies midway between two points of the transform then it will appear to have a lower magnitude.

A8.1.2 Window Coefficients

We will not discuss this in any detail here, and will merely note that the coefficients for various cosine windows are given in the program listing for the Matlab function `cosinewindow` in §A8.2.2. For the graphs presented in chapter 10, we used a minimum 4-term Blackman-Harris window, but we modified the co-efficients slightly, to cause the sidelobe heights to drop off at 18dB/octave instead of at 6dB/octave. The coefficient 0.01168 was replaced by 0.01174 so that the coefficients summed to zero. The overall performance was virtually unaltered, and the exercise had no real validity other than to verify an obscure point arising from a detailed reading of (Harris, 1978).

A8.2 MatLab's DFT Function

This Appendix is simply an *aide memoire* in which we confirm the nature of MatLab's Fourier transform function.

A8.2.1 Distribution of Coefficients

The distribution of the constants in a Fourier transform pair is not unique, so it can be helpful to verify the performance of the transforms that are used for a spectral analysis. MatLab provides a discrete Fourier transform (DFT) function – called FFT – which maps $f(nT) \leftrightarrow F(m\Omega)$ such that $f(1) \leftrightarrow F(N\delta(0))$, where N is the number of points in the transform. For example, the following vectors

$$[1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1] \leftrightarrow [8 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

form a DFT pair. The implication is that the frequency spectrum (apart from the zero-frequency point) must be divided by $\frac{1}{2}N$ to correctly represent the signal voltage at each frequency; e.g. consider the sine wave

$$[0 \ \frac{1}{\sqrt{2}} \ 1 \ \frac{1}{\sqrt{2}} \ 0 \ -\frac{1}{\sqrt{2}} \ -1 \ -\frac{1}{\sqrt{2}}]$$

The DFT of this, in MatLab, is

$$[0 \ 4 \ 0 \ 0 \ 0 \ 0 \ 0 \ 4]$$

and dividing by 4 gives the magnitude of the sine wave as 1.

A8.2.2 Signal Power

We can also run a check on the signal power. If we square the terms in the frequency spectrum and divide by two, this represents the mean-square voltage (which is proportional to power) at each frequency. Summing these terms from 0 to $N/2$ gives the total power. In the above example, the frequency spectrum is

$$[0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1]$$

and the total power is $\frac{1}{2}$. The sum of the squares of the values in the time waveform is

$$0 + \frac{1}{2} + 1 + \frac{1}{2} + 0 + \frac{1}{2} + 1 + \frac{1}{2} = 4$$

and dividing this by N gives the mean square voltage of the waveform, which is $\frac{1}{2}$ and agrees with the previous calculation.

A8.2.3 Sample Numbering

Lastly, we recap. on the sample counting. We are using a 131072-point transform, with the first sample, $F(0)$ corresponding to d.c. $F(65536)$ corresponds to 250kHz. $F(65537)$ is a copy of $F(65535)$ and so on, up to $F(131071)$ which is a copy of $F(1)$. MatLab confuses the issue by requiring the index to run from 1 to 131072. In addition, there are $N + 1$ points on an x -axis that is configured to run from 0–250kHz.

A8.3 References

Harris, F. J. (1978). On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. *Proc. IEEE* 66(1), 51-83.

A9 Program Listings on CD-ROM

A9.1 MatLab Programs

The reader will need to be familiar with MatLab in order to fully understand the programs listed here. MatLab uses a conventional ‘third-generation’ algorithmic language, but with several features that distinguish it from similar languages. The most important of these is that variables are treated as arrays or matrices, and there is therefore a distinction between, for example, vector or matrix multiplication and element-by-element multiplication of those entities. In addition, it should be realised that MatLab has been used much as a ‘pocket calculator’ would be, and the graph-plotting routines listed here have minimal documentation – these listings are not ‘elegant’ examples of MatLab programs!

A9.1.1 Propagation

Figure number	Generated files	MatLab program	Subroutines
Figure 2-4	skin-1.tif	PHD_skin1	skindepth.m
Figure 2-1	loop-1.tif, loop-2.tif	PHDmagloop1.m	
Figure 2-3	loop-3.tif		
Figure 2-5	loop-4.tif, loop-5.tif		air.m
Figure 2-6	loop-6.tif		plotphase1.m
Figure 2-7	loop-7.tif	PHDmagloop2.m	PQ-integral.m
Figure 2-8	loop-12.tif, loop-11.tif		rock1.m
Figure A2-3	loop-9.tif		rock2.m
Figure 2-9	loop-10.tif		rock3.m
Figure 2-13	loop-8.tif		sommerfeld.m
Figure A2-4	QvD-1.tif, QvD-2.tif		
Figure A2-5	QvD-3.tif, QvD-4.tif	PHD_QvD.m,	
Figure 2-14	QvD-5.tif, QvD-6.tif	PHD_QvD_data.mat,	
Figure 2-15	QvD-7.tif, QvD-8.tif	PHD_QvD_data.m	
Figure 2-16	QvD-11.tif, QvD-12.tif		
Figure A2-6	PvD-1.tif, PvD-2.tif	PHD_PvD.m,	
Figure A2-7	PvD-3.tif, PvD-4.tif	PHD_PvD_data.mat,	
Figure 2-17	PvD-5.tif, PvD-6.tif	PHD_PvD_data.m	
Figure 2-18	opt1.tif, opt2.tif	PHD_QvF1.m	
Figure 2-19	opt3.tif	PHD_QvF2.m	

A9.1.2 Atmospheric Noise

Figure number	Generated files	MatLab program	Subroutines
Figure 7-4	loop3-10.tif	PHDmagloop3.m	As above
Figure 7-3	opt-6.tif, opt-5.tif	PHD_QvF3.m	
Figure 7-5	opt-3.tif, opt-4.tif		

Notes: file opt-3.tif is not the same file referred to in §A9.1.1. These programs use some of the subroutines from the previous section.

A9.1.3 Sequence Design for System Identification

Figure number	Generated files	MatLab program	Subroutines
Figure 9-2	iseq-1.tif		
Figure 9-3	iseq-2.tif	PHDinvseq1.m	pconv.m
Figure 9-5	iseq-3.tif		timereverse.m
Figure 9-7	iseq-4.tif		pcorr.m
Figure 9-8	iseq-6.tif, iseq-5.tif		seq_snr.m
Figure 9-9	iseq-8.tif, iseq-7.tif		

A9.1.4 Spectrum of M+1 Sequence

Figure number	Generated files	MatLab program	Subroutines
Figure 8-2	seq-2.tif, seq-1.tif	PHD_seqpower.m	

A9.1.5 Channel Sounder Results

Figure number	Generated files	MatLab program	Subroutines
Figure 10-3 to	various	PHDchsound1.m	chs4p3_sub.m
Figure 10-16			*.mat, ch3.m

Note: the *.mat files that comprise the results data were created by running chs3.m on the raw data taken from the channel sounder's PCMCIA hard disc, and were then manually saved using Matlab's 'save' command. Neither the raw data nor documentation on how to run chs3.m is included on the CD-ROM.

A9.1.6 Window Functions

Figure number	Generated files	MatLab program	Subroutines
Figure A8-1	window-1.tif, window2.tif	window.m	cosinewindow.m gencoswin.m

A9.2 EPLD Programs

File name	Subject	Compile using...
Chspg1v1.txt	listing for PG1 (see §A7.4.1)	PLPL
Chspg2v3.txt	listing for PG2 (see §A7.4.2)	PLPL
Chspg3v3.abl	listing for PG3 (see §A7.4.3)	Scenario

A9.3 Spreadsheet Programs

Text Reference	Excel program
Table 5-5, 5-6 & discussion in §5.8.3	Antenna3.xls
Figure 6-4 & discussion in §6.2.6; Figure 1-2 & Table 1-1	Noise-5.xls



Submitted thesis: A CD-ROM is attached to the inside back cover.

**The CD may be absent from other copies, please enquire to the author
(see address in appendix A1.6)**

**THESIS
CONTAINS
CD/DVD**