



OXFORD

the
Photomultiplier
A G
WRIGHT Handbook

THE PHOTOMULTIPLIER HANDBOOK

The Photomultiplier Handbook

A. G. Wright

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Tony Wright 2017

The moral rights of the author have been asserted

First Edition published in 2017

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2017933575

ISBN 978-0-19-956509-2

DOI: 10.1093/oso/9780199565092.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Dedicated in memory of John Barton.

Contents

Preface	xix
Figure credits	xxi
1 Why photomultipliers?	1
1.1 Aspects of light detection	2
1.1.1 Introduction	2
1.1.2 Brief history of PMTs	3
1.1.3 The statistical case for PMTs	4
1.1.4 DC detection with a PMT	7
1.1.5 Detection of single photoelectrons	9
1.1.6 Detection of multi-photoelectron signals	10
1.1.7 Summary of PMT key attributes	11
1.2 Other light detectors	12
1.2.1 Silicon PIN diodes	12
1.2.2 APDs	12
1.2.3 SiPMs	14
1.2.4 Summary of silicon key attributes	16
1.2.5 Visible-light photon counters	17
1.2.6 CCDs	17
1.2.7 Hybrid photodetectors and APDs	17
1.3 Pulse height resolution	18
1.4 Position resolution	21
1.5 Signal-to-background considerations	21
2 Photocathodes	24
2.1 Introduction	25
2.1.1 Solid angles	26
2.2 Fundamentals of photosensitivity	27
2.2.1 The electromagnetic spectrum	27
2.2.2 Photoelectric phenomena	28
2.2.3 Photoelectron energy distribution	29
2.3 Spectral radiation and photometric units	30
2.3.1 Spectral radiant energy	30
2.3.2 Photometric units and standards	30
2.3.3 Filter measurements	33
2.3.4 Calibration laboratories and absolute radiation standards	35
2.3.5 Trap detectors	35

2.4	Optical properties of end window photocathodes	37
2.4.1	Refraction in dielectrics	37
2.4.2	Reflection from dielectrics	40
2.4.3	Photocathode optical properties	42
2.4.4	Techniques for measuring photocathode reflectance	43
2.4.5	Limits to reliable measurement	44
2.4.6	Determination of transmission and absorptance	45
2.5	Spectral sensitivity of photocathodes	50
2.5.1	Three foremost photocathodes	50
2.5.2	Photocathode response to scintillator light	52
2.5.3	Solar blind photocathodes	54
2.5.4	Infrared photocathodes	56
2.5.5	High-temperature photocathodes	58
2.6	Photocathode geometries	58
2.6.1	Range of end window PMTs	58
2.6.2	Range of side window PMTs	60
2.6.3	Window materials	61
2.7	Photocathode resistivity	62
2.8	QE temperature dependence	65
2.8.1	Laboratory and commercial applications, $-30\text{ }^{\circ}\text{C}$ to $+60\text{ }^{\circ}\text{C}$	65
2.8.2	Ultra-low temperature performance, $-273\text{ }^{\circ}\text{C}$ to $-30\text{ }^{\circ}\text{C}$	67
2.8.3	High-temperature operation, $+60\text{ }^{\circ}\text{C}$ to $+200\text{ }^{\circ}\text{C}$	68
2.9	Dark count temperature sensitivity	69
2.10	Critical photocathode properties	69
2.10.1	Photocathode uniformity	69
2.10.2	Ionizing radiation	72
2.11	Summary of factors affecting photocathode response	73
3	The optical interface to PMTs	78
3.1	Introduction	79
3.2	Light attenuation and reflection coefficients	80
3.2.1	Light attenuation length, γ	80
3.2.2	Reflectance coefficients, ρ	81
3.2.3	Optical coupling of scintillators	83
3.3	Point light source	84
3.4	Interfacing to small volume scintillators	86
3.4.1	Interfacing to scintillators of high Z	88
3.4.2	Interfacing to scintillators of low Z	90
3.5	Light concentrators	90
3.5.1	Light concentrators: Theory	90
3.5.2	Light concentrators: Winston cone	91
3.5.3	Integrating spheres	92

3.6	Scintillators with diffuse reflecting walls	93
3.7	Light output from scintillator slabs	95
3.7.1	Escape cones	95
3.8	Light guides	97
3.8.1	Hollow guides with specular reflecting walls	98
3.8.2	Solid guides based on TIR	98
3.8.3	Adiabatic light guides	99
3.9	Light detection efficiency for scintillators	101
3.9.1	PMT viewing a large-area face	101
3.9.2	PMT viewing small-area face	103
3.9.3	Photoelectron yield for scintillation counters	105
3.9.4	Monte Carlo simulations	107
3.10	Uniformity of response in large-area detectors	107
3.11	WLS light guides	109
3.12	Optical fibres	110
3.12.1	Application of fibres to scintillating tiles	112
3.12.2	Use of fibres with large-area plastic scintillator slabs	114
3.13	Techniques for enhancing light detection	115
3.13.1	Anti-reflection coatings	115
3.13.2	Sandblasting	115
3.13.3	Theory of light recycling	117
3.13.4	Internal prismatic window	119
3.13.5	External optical enhancement devices	120
3.13.6	WLS coatings	123
4	Statistical processes	128
4.1	Introduction	129
4.1.1	Single- and multi-photon light sources	129
4.2	Binomial, Poisson, and normal distributions	131
4.3	Mean and variance	132
4.4	Folding of probability density functions	136
4.5	Moments	139
4.5.1	Moments about the origin	139
4.5.2	Central moments	140
4.5.3	Factorial moments	141
4.5.4	Exercise in photon counting	142
4.6	Probability generating functions	145
4.7	Sum of chance integers	147
4.8	Moment generating functions	149
4.9	Folding with probability generating functions	151
4.10	Cascaded processes	153
4.10.1	Generalized two-stage cascading	153
4.10.2	Light detection statistics	154
4.10.3	Multiplier dispersion	155
4.10.4	Pdf at the k^{th} dynode	157

4.11	Shot noise in DC detection	161
4.12	Noise in multi-photon excitation	163
4.13	Continuous SER distributions	166
4.14	Multi-photoelectron pdfs	169
4.15	MPHD for an empirical SER	173
4.16	Arrival time statistics	175
4.16.1	Exponential time signatures	175
4.16.2	First two moments for R fixed	177
4.16.3	First two moments for R randomized	178
4.17	Determination of afterpulse rates	180
4.17.1	Afterpulses: Method I	181
4.17.2	Afterpulses: Method II	182
4.18	Scaling and dead time	183
4.18.1	Interval distribution	184
4.18.2	The p-fold time interval	184
4.19	Dead time	188
4.19.1	Sources of dead time	188
4.19.2	Counting losses attributed to dead time	190
4.19.3	Type I, paralysable counters	190
4.19.4	Type II, non-paralysable counters	191
4.19.5	Dead time and variance	193
4.19.6	Determination of dead time	194
4.19.7	Suppressing correlated signals	197
5	Secondary emission and gain	202
5.1	Introduction	203
5.2	Gain generation with discrete dynodes	204
5.3	Single-photoelectron noise spectra	208
5.4	Secondary-electron emission	210
5.4.1	Classical description	210
5.4.2	Secondary-electron yield	214
5.4.3	Theory and measurement	215
5.4.4	Secondary-electron statistics	216
5.4.5	Temperature dependence of gain	218
5.5	Sources of undersized signals	219
5.5.1	Inelastic scattering	219
5.5.2	Gain non-uniformity	222
5.5.3	Events initiated by d_1	224
5.5.4	Other sources of undersized signals	224
5.6	Discrete dynodes	225
5.6.1	Venetian blind dynodes and skipping	226
5.6.2	Coarse mesh dynodes	229

5.6.3	Fine-mesh dynodes	230
5.6.4	Metal-channel dynodes	230
5.6.5	Linear and circular focussed dynodes	232
5.6.6	Box-and-grid dynodes	234
5.6.7	Transmission dynodes	234
5.7	Continuous multipliers	237
5.7.1	CEMs	237
5.7.2	MCPs	239
5.7.3	Gain–voltage considerations in MCPs	240
5.8	Multiplier and PMT gain measurement	241
5.8.1	Measurement of G: First method	242
5.8.2	Measurement of G: Second method	243
5.8.3	Measurement of G: Third method	243
5.8.4	Measurement of $\langle g \rangle$: First method	245
5.8.5	Measurement of $\langle g \rangle$: Second method	247
5.8.6	Measurement of $\langle g \rangle$: Third method	248
5.8.7	Measurement of $\langle g \rangle$: Fourth method	249
5.9	Quick determination of $\langle g \rangle$	249
5.9.1	Quick determination of $\langle g \rangle$: First method	251
5.9.2	Quick determination of $\langle g \rangle$: Second method	251
5.10	Gain–voltage characteristics	252
5.11	Conclusions	253
6	PMT background	257
6.1	Introduction	258
6.2	Dark counts and dark current	259
6.2.1	Subdivision of a dark count distribution	260
6.3	Reconciliation of dark current and dark counts	261
6.4	S/B in DC applications	264
6.5	Light generation within a PMT	265
6.6	Sources of background	266
6.6.1	Region A: Undersized pulses	267
6.6.2	Regions B and C: Single- and multi-electron background	267
6.6.3	Dependence on temperature	267
6.6.4	Ionizing radiation	269
6.7	Gamma background	271
6.7.1	Radionuclides within PMTs	274
6.8	Cosmic ray muons	276
6.8.1	Cerenkov emission in a PMT window	277
6.8.2	PMT orientation	278
6.9	Means for reducing background	279
6.10	Summary	279

7 Measurement of low light flux	282
7.1 The physical nature of light detection	283
7.1.1 Noise and background	283
7.2 Measurement modes	284
7.2.1 Photon counting	284
7.2.2 DC detection	284
7.2.3 Shot noise power detection	286
7.3 Detection methods and weighting	288
7.3.1 Detection techniques	289
7.4 Difference between two count rates	291
7.4.1 Efficient time allocation	291
7.5 Signal recovery instrumentation	293
7.5.1 Boxcar integrators	293
7.5.2 Lock-in detection	294
7.5.3 Synchronous signal averaging	296
7.5.4 Signal recovery simulation	297
7.5.5 Subdividing time intervals	298
7.5.6 Upper and lower thresholds	300
7.5.7 Signal magnitudes: Photon counting	301
7.5.8 Signal magnitudes: DC detection	302
7.5.9 Excess noise factor	303
7.5.10 Ratemeter operation	304
7.6 PMT selection criteria	306
7.6.1 Optimal gain for DC detection	307
7.6.2 Optimal HV for photon counting	308
7.6.3 MCA plateau characteristic	311
7.6.4 Critical PMT parameters	312
7.6.5 Discriminator threshold	316
7.7 Methods involving DMMs and electrometers	317
7.7.1 Measurements with a DMM	317
7.7.2 Electrometers	318
7.7.3 Voltage dividers	319
8 Timing with PMTs	321
8.1 Electron motion in electric fields	322
8.1.1 Introduction	322
8.1.2 Electron motion in a uniform electric field	328
8.1.3 Electron motion in non-uniform electric fields	329
8.1.4 Laws of motion applied to PMTs	330
8.1.5 Spot size	331
8.2 Evolution of fast linear focussed PMTs	333
8.2.1 The PMT front end	333
8.2.2 Contribution to time dispersion from dynodes	337

8.2.3	Transit-time effects at the PMT back end	339
8.2.4	Ultra-fast PMTs	341
8.3	PMT output signal	341
8.3.1	Signature of the anode signal	341
8.3.2	Simulation of photoelectron time intervals	343
8.3.3	Simulation of multiplier noise and jitter	345
8.3.4	Oscilloscope traces for NaI(Tl) signals	347
8.3.5	Scintillator rise and fall time	350
8.4	Timing	351
8.4.1	Single-electron time response	351
8.4.2	Transit time, T	353
8.4.3	Effect of amplitude fluctuations on jitter	356
8.4.4	Effect of charge sensitivity on jitter	357
8.4.5	Contribution to jitter from noise	357
8.5	Zero crossing and CF methods	358
8.5.1	Charge threshold for triggering	361
8.5.2	CF technique	363
8.6	Measurement schematics	365
8.7	Multiple statistical sources	369
8.7.1	Single-photoelectron output signatures	369
8.7.2	Photoelectron timing variance	370
8.8	Photon arrival statistics	373
8.9	Timing measurements	376
8.10	Timing summary	378
9	Linear performance	383
9.1	Introduction	384
9.1.1	Modes of operation	385
9.1.2	Rate effect	387
9.1.3	Causes of non-linearity	387
9.1.4	Dynode materials	387
9.2	Dynamic range	388
9.2.1	Achievable dynamic range: Analogue	388
9.2.2	Achievable dynamic range: Pulse mode	389
9.2.3	Onset of DC non-linearity	391
9.2.4	Onset of pulsed non-linearity	392
9.2.5	Bootstrap method	394
9.3	Theoretical considerations	394
9.4	Advanced linearity investigations	398
9.4.1	Single-step method	399
9.4.2	Ratiometric method	402
9.4.3	Shot noise method	403
9.5	Correcting non-linearity	404
9.6	Non-linearity in inorganic crystals	405
9.7	Summary	407

10 Collection and counting efficiency	410
10.1 Introduction	411
10.2 Lost photoelectrons	412
10.2.1 Front end design	412
10.2.2 Optimal operating conditions	413
10.3 Collection and counting efficiency	414
10.3.1 Collection efficiency, F	414
10.3.2 Anode detection efficiency, ϵ	415
10.4 Measurement techniques	415
10.4.1 Attenuation using filters	415
10.4.2 Count rate determination	416
10.5 Determination of F	418
10.5.1 Based on d_1 gain	418
10.5.2 Shot noise	422
10.5.3 The bi-photon field	424
10.5.4 Calibrated lamp	425
10.5.5 Cerenkov light source	425
10.5.6 Use of a power meter	426
10.5.7 Calibrated silicon photodiodes	427
10.6 Summary and conclusions	428
11 Signal-induced background	431
11.1 Introduction	432
11.2 Timing instrumentation	432
11.2.1 Amplifier discriminators	432
11.2.2 Configurations for afterpulse measurements	433
11.2.3 Stop–start instrumentation	434
11.3 Prepulses and late pulses	435
11.4 Detector response function	439
11.5 Afterpulses	440
11.5.1 Early measurements	440
11.5.2 Theory of afterpulse generation	442
11.5.3 Contributions to understanding afterpulses	445
11.5.4 Afterpulses in photon counting	447
11.5.5 Exposure of PMTs to helium	448
11.6 Exposure to bright light	449
11.6.1 Settling time: Medium term	449
11.6.2 Settling time: Long term	449
11.7 Summary and conclusions	452
12 Environmental considerations	455
12.1 Introduction	456
12.1.1 Performance in weak magnetic fields	456
12.1.2 Performance in high magnetic fields	458

12.1.3	The effect of electric fields	458
12.1.4	PMT enclosures, modules, and housings	461
12.1.5	Housings for scintillators	463
12.1.6	Light leaks and electric fields	464
12.2	Operation in harsh environments	465
12.2.1	Shock and vibration	465
12.2.2	Operation at high temperature	467
13	Voltage dividers	471
13.1	Introduction	472
13.2	Gain–voltage relationships	473
13.2.1	Resistive dividers	473
13.2.2	Dividers with zener diodes	475
13.3	Analysis of resistive dividers	477
13.4	Resistive divider operation	482
13.5	Inclusion of zener diodes	484
13.6	Active voltage dividers	488
13.6.1	Fully active N-type analysis	489
13.6.2	Fully active P-type analysis	490
13.6.3	Fully active circuits	493
13.6.4	Gain linearity	495
13.6.5	Partially active circuits	496
13.6.6	Cockcroft–Walton dividers	497
13.7	Voltage dividers for pulsed operation	499
13.7.1	Decoupling capacitors	499
13.7.2	Pin connections	501
13.7.3	Decoupling options	501
13.7.4	Switch-on transients	502
13.8	Methods and circuits	503
13.8.1	Power supply polarity and ripple	503
13.8.2	Wiring options	506
13.8.3	Dynode signals	507
13.8.4	Shorting dynodes	509
13.8.5	Single cable dividers	510
13.8.6	Gain adjustment	511
13.8.7	Equivalent circuit at high frequencies	511
13.8.8	Wiring practice	512
13.9	Gating	515
13.9.1	Electro-optical and mechanical gating	517
13.9.2	HV switching	517
13.9.3	Photocathode gating	518
13.9.4	Dynode gating	518
13.9.5	Focus-electrode gating	521
13.9.6	Circuit realization	522
13.9.7	Pulse generators and shielding	522

14 Electronics for PMTs	527
14.1 Introduction	528
14.2 Charge output transducers	530
14.3 The role of a preamplifier	531
14.3.1 Interfacing and signal transmission	532
14.3.2 Grounding and shielding	534
14.4 Circuit protection and limiting	534
14.5 Modular electronics	535
14.6 PMT equivalent circuit	538
14.6.1 PMT signal bandwidth	541
14.7 Multiple PMT outputs	543
14.8 Passive circuit analysis	544
14.8.1 Anode DC-coupled R C network	545
14.8.2 Anode AC-coupled R C network	547
14.9 Signal-shaping preamplifiers	549
14.9.1 Charge-sensitive preamplifiers	550
14.9.2 Commercial charge-sensitive amplifiers	552
14.9.3 Pole-zero cancellation	554
14.9.4 Preamplifier calibration	555
14.10 Signal-shaping main amplifiers	555
14.10.1 Signal-shaping theory	555
14.10.2 Detector noise	559
14.10.3 Modular detection equipment	561
14.11 Fast amplifiers	562
14.11.1 Discrete-component amplifiers	562
14.11.2 Amplifiers incorporating microwave transistors	564
14.11.3 Fast bench-top and NIM preamplifiers	566
14.11.4 Voltage- and current-feedback OP-AMPS	566
14.11.5 Transimpedance amplifier circuits using CFAs	569
14.11.6 Logarithmic amplifiers	571
14.12 Pulsed-light sources	572
14.13 Optimizing PMT performance	573
14.14 DSP	574
Appendix A PMT output distributions	581
A.1 General considerations	581
A.2 Schematics and methods for measuring distributions	583
A.3 Integral pulse height distributions	583
A.4 Differential pulse height distributions	583
A.5 Plateau characteristic	584
A.6 Changing variables	585
A.7 Transformations of an analytic function	588
A.8 Transformation of experimental data	591

A.9	Single-photoelectron distributions	592
A.10	Probability distributions in high temperature NaI(Tl) applications	597
A.11	Conclusions	600
Appendix B Light emission by the Cerenkov effect		603
B.1	Introduction	603
B.2	Threshold conditions for the Compton effect	604
B.3	Formulations for Cerenkov emission	606
B.4	The yield from MIPs	608
Appendix C Abbreviations		611
Index		615

Preface

It is usual to indicate the type of reader the author has in mind. This book is aimed primarily at those who use, or are about to use, vacuum photomultipliers (PMTs). The aim is to provide a deeper understanding of PMT behaviour as a means for optimizing performance. Wherever appropriate I have drawn on my experience as a user and manufacturer of PMTs. My objective in writing this book is to fill perceived gaps in the literature. For example, the important topic of the optical interface to PMTs has hitherto received little attention.

Major manufacturers of light detectors, notably EMI, provided ‘Application Notes’ covering a range of technical topics relating to PMTs. Also, PMT manufacturers’ handbooks, refreshingly unbiased by commercial considerations, have been freely issued over the years, namely,

- RCA (*Photomultiplier Manual*, 1970)—available online
- Philips (*Photomultiplicateurs*, 1981)
- Hamamatsu (*Photomultiplier Tubes, Basics and Applications*, 2006)—available online
- Philips (*Photomultiplier Tubes, Principles and Applications*, 1994)
- *Photomultiplicateurs*, written by G Piétri, in French, reached a wider audience following an update and translation released in 1994.

It is perhaps worth noting the name changes of the major manufacturers, as companies have been bought or merged. The evolution of names is as follows:

- Philips → Photonis → now based in China
- EMI → Thorn EMI → Electron Tubes → ET Enterprises
- RCA → Burle → ceased trading in PMTs

Philips marketed and branded products under different titles: Philips, Mullard, Valvo, and Amperex. When a specific product type is mentioned in the present text, the name of the manufacturer at the time of reporting is used.

This book should also be of interest to scientists and engineers involved in PMT design and manufacture. I have held the opinion throughout my 40 years in the industry that those who produce new or modified devices should critically evaluate their performance. This also applies to designers of electronics, for whom dealing with the equivalent circuit of a current generator may initially appear

arcane. There is sometimes reluctance to investigate PMT behaviour occurring at a stage prior to the output, because performance is affected by operations before the signal reaches the anode. PMTs have been commercially available since the 1950s, but optimal or even satisfactory performance still eludes many users. This is not surprising, since we are dealing with devices of sufficient gain to detect single photons and with a bandwidth of the order of a gigahertz—clearly, a product with such performance demands expertise and careful treatment.

PMTs are intriguing scientific transducers that have proved worthy of study in their own right. They embody the physics of photoemission, electron-optics, secondary emission, optics, and signal processing, making them the preferred light detectors for a significant range of commercial instrumentation and for academic investigations. The principal advances in the evolution of PMTs inevitably trace back to a scientific requirement, rather than a commercial one. Hemispherical PMTs (see Fig. 2.23), for example, stemmed from proton-decay investigations requiring the capture of light over large solid angles. Following their first introduction in the 1970s, many massive experiments conducted under water, and in ice, have produced outstanding advances in our understanding of astrophysical phenomena. Although perhaps 50,000 PMTs of this type have been produced, virtually none has found commercial application. The majority of scientific publications concerning PMTs are to be found in the following journals: *Nuclear Instruments and Methods in Physics Research A*, *IEEE Transactions on Nuclear Science*, *Applied Optics*, *Review of Scientific Instruments*, and *Journal of Physics D* and *E*. The first two mentioned are favoured for publishing the proceedings of international conferences, such as: ‘New Developments in Photodetection’ and the ‘IEEE Nuclear Science Symposia’. These conference publications are highly recommended as a source of current information on light detectors.

Light generation and detection are statistical processes. The family of detectors that provide gain through electron multiplication are also statistical in their operation. Hence, the choice of an appropriate detector for a particular application is strongly influenced by statistical arguments. The first chapter introduces the range of light detectors and compares their performance, essentially on statistical grounds. The formulations are stated rather than derived at this stage but they are fully developed in Chapter 4.

I have been fortunate to have enjoyed constant encouragement, the use of test facilities, and literature from two managing directors of Thorn EMI and ET Enterprises, namely, J P Frederiksen and R M McAlpine, respectively, and also L Ludlum, Vice President, ADIT/Electron Tubes. I also wish to acknowledge A C Bach, A J Cormack, P J Cook, and C Wade for technical input and suggestions on presentation and subject coverage. I am especially indebted to my wife, Mary, who patiently withheld the awkward question ‘When are you going to finish?’

*Amersham
June 2017*

A. G. WRIGHT

Figure credits

Chapter 2

Fig. 2.10(a). Jones, D. P. (1976). Photomultiplier sensitivity variation with angle of incidence on the photocathode. <i>Appl. Opt.</i> , 15, No. 4, 910–14.	44
Fig. 2.10(b). Moorhead and Tanner (1996). Optical properties of an EMI K ₂ CsSb bialkali photocathode. <i>Nucl. Instr. and Meth. in Phys. Res. A</i> , 378, 162–70. Reprinted from Elsevier Science ©1996.	44
Figs 2.11 to 2.14. Moorhead and Tanner (1996). Optical properties of an EMI K ₂ CsSb bialkali photocathode. <i>Nucl. Instr. and Meth. in Phys. Res. A</i> , 378, 162–70. Reprinted from Elsevier Science ©1996.	46–8
Fig. 2.15. Hallensleben <i>et al.</i> (2000). Optical constants for the S20 photocathode, and their application to increasing photomultiplier quantum efficiency. <i>Optics Communications</i> , 180, 89–102. Reprinted from Elsevier Science ©2000.	49

Chapter 3

Figs 3.9 to 3.11. Welford, W. T. and Winston, R. (1978). <i>The optics of non-imaging concentrators. Light and solar energy</i> . Academic Press, New York. Reprinted from Academic Press ©1978.	90–2
Fig. 3.26. Aota, S. <i>et al.</i> (1999). Mass production of tile/fibre units for the CDF plug upgrade EM calorimeter. <i>Nucl. Instr. and Meth. in Phys. Res. A</i> 420, 48–61. Reprinted from Elsevier Science ©1999.	113
Fig. 3.27. Artikov, A. <i>et al.</i> (2006). New generation large area muon scintillation counters with wavelength shifter fibre readout for CDF II. <i>Physics of particles and nuclear letters</i> , 3, Issue 3, 188–200. Reprinted from Springer ©2006.	114
Fig. 3.34. Gunter, W. D., Grant, G. R. and Shaw, S. A. (1970). Optical devices to increase photocathode quantum efficiency. <i>Appl. Opt.</i> , 9, No. 2, 251–7. Reprinted with permission from OSA ©1970.	122
Table 3.2. Courtesy of J Mc Millan, private communication.	83

Chapter 4

Figs 4.14 and 4.15. Wright A. G. (2005). The statistics of multi-photoelectron pulse height distributions. <i>Nucl. Instr. and Meth. in Phys. Res. A</i> , 579, 967–72. Reprinted from Elsevier Science ©2005.	174, 177
Fig. 4.16. Arrival time statistics. Kelbert, M., Sazonov, I. and Wright, A. G. (2006). Exact expression for the variance of the photon emission process in scintillation counters. <i>Nucl. Instr. and Meth. in Phys. Res. A</i> , 564, 185–9. Reprinted from Elsevier Science ©2006.	179

Chapter 5

Fig. 5.10. Simon, R. E. and Williams, B. F. (1968). Secondary-electron emission. <i>IEEE Trans NS15</i> , 166–70. Reprinted with permission from IEEE. ©1968.	215
---	-----

xxii Figure credits

- Fig. 5.12. Sommer, A. (1972). Bialkali (K_2CsSb) photocathodes as a high gain secondary electron emitter. *J. Appl. Phys.*, 43, No. 5, 2479–2480. Reprinted with permission from AIP© (1972). 217
- Fig. 5.23. Kuroda, K., Sillou, D., and Takeutchi, F. (1981). New type of position sensitive photomultiplier. *Rev. Sci. Instrum.*, 52(3), 337–346. Reprinted with permission from AIP© (1981). 229
- Fig. 5.24. Agoritsas, V., Kuroda, K. and Nemoz, Ch. (1989). A new type of position-sensitive electron multiplier. *Nucl. Instr. and Meth. in Phys. Res. A*, 277, 237–241. Order number 3620250895810. Reprinted from Elsevier Science ©1989. 229
- Fig. 5.26. Chirikov-Zorin, I., Fedorko, I., Menzione, A., Pikna, M., Sýkora, I., and Tokář, S. (2001). Method for precise analysis of the metal package photomultiplier single photoelectron spectra. *Nucl. Instr. and Meth. in Phys. Res. A*, 456, 310–324. Reprinted from Elsevier Science ©2001. 231

Chapter 6

- Table 6.1. Viehmann, W, et al. (1975). Photomultiplier window materials under electron irradiation: fluorescence and phosphorescence. *Appl. Opt.*, 14, No. 9. 2104–15. Reprinted with permission from OSA ©1975. 271

Chapter 10

- Figs 10.5 to 10.11 and equations (10.6) and (10.7). Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, 49, 2059–65. Reprinted with permission from OSA ©1970. 419–23

Chapter 14

- Fig. 14.24. Blanch, O. et al. (1999). Performance of a fast low noise front-end preamplifier for the MAGIC imaging Cerenkov telescope. *IEEE Trans. NS.* 46, 800–5. Reprinted with permission from IEEE. ©1968. 565
- Fig. 14.25, Fig. 14.26, and Fig. 14.27. Giachero, A., Gotti, C., Maino, M., and Pessina, G. (2011). Current feedback operational amplifiers as fast charge sensitive preamplifiers for photomultiplier read out. *J Inst.* 6, P05004, 1–19. Reprinted with permission from IOP ©2011. 567, 568, 570

1

Why photomultipliers?

1.1 Aspects of light detection

1.1.1 Introduction

Photon detectors transform information, carried by light, to an electrical analogue. A light signal primarily contains information on: time relative to some initiating event; intensity in terms of a number of photons; position of a detected signal on the photocathode; and wavelength. Not all this information is necessarily sought by a user, and some detectors are superior to others concerning a particular performance attribute. Intensity may refer to the number of photons per event, or per unit time, with consideration for the detector area. Photon rates may be constant with time (DC detection), slowly varying, or transient in nature. The time response of a detector is generally specified in terms of some property of the pulse shape, such as its rise time; or the response may be described in terms of bandwidth. All detectors are sensitive to the wavelength of a signal but without the capability for its direct determination: this demands additional instrumentation, such as a monochromator.

The process of light detection is a quantum mechanical one of absorbing the energy of a photon to produce a free electron. Given the generation of sufficient photoelectrons per unit time, it is possible to quantify this charge in terms of a current or voltage to provide a measure of the light input. The ability to detect a light signal depends broadly on two considerations. The first is the sensitivity of the photocathode, in terms of quantum efficiency (QE), η : that is, the ratio of the number of photoelectrons produced to the number of incident photons and second, the level of background. QE is always less than unity, although in certain solid state detectors the QE approaches unity over a limited wavelength band. Other detectors, such as photomultipliers (PMTs), are often used at infrared wavelengths, where QEs are very low. It is, however, still feasible to make measurements under these circumstances since PMTs are capable of detecting single photoelectrons, by virtue of their high and relatively noiseless gain. Until the 1970s PMTs were the only commercially available devices with this capability, but now solid state PMTs, SiPMs, with intrinsic gain are readily available.

A brief introduction to the terminology is appropriate at the outset, to remove possible confusion. The following terms are encountered throughout the scientific literature: signal current and signal counts, where the meanings are obvious. Associated with these two variables we have dark current and dark counts, or equivalently, background current, and background counts. The words dark and background apply to the output from a detector in the absence of light input but the terms actually refer to signals, albeit unwanted ones. Noise is reserved for the description of statistical aspects of light, such as shot noise; it is not a signal but rather a fluctuation within an existing one. ‘Noise in signal’ and ‘noise in background’ are the appropriate terms for the shot noise inherent in the current produced by a light signal, and that produced by a background signal, respectively.

Throughout this book background and noise will always refer to different aspects of performance, as described here.

Light detectors may be divided into two broad categories: those that provide an image and those that are non-imaging. Further subcategories are:

- vacuum devices, such as PMTs, microchannel plate PMTs (MCPMTs), and hybrid PMTs, in which a discrete multiplier is replaced by a solid state diode
- solid state devices, such as PIN diodes, avalanche photodiodes (APDs); pixelated silicon PMs (SiPMs), and charge-coupled devices (CCDs)

Throughout the book, the acronym PMT refers to traditional vacuum PMTs.

1.1.2 Brief history of PMTs

The invention of the PMT in 1936 is usually credited to the RCA Corporation in the USA. However, Lubsandorzhiev (2006) makes a case in favour of L A Kubetsky, a Russian, who produced a PMT in 1933–4 with an S1 photocathode, consequently predating the work at RCA. This is arguable, but what cannot be disputed is the leading role of RCA in the development of commercially available product, starting in 1934. PMTs evolved from studies in electron secondary emission at the turn of the twentieth century. It was observed, for certain metals, that the process of secondary electron emission generated more electrons than primary emission did—hence a means of amplification. RCA produced a single-stage PMT in 1935, with a gain of 8 (Engstrom 1980). A multistage magnetically focussed PMT followed in 1936 from researchers who were to become the leaders in the PMT industry—Zworykin, Morton, and Malter (1936). This PMT was not commercialized because of the difficult operating conditions imposed by the need for both electrostatic and magnetic focussing. The first commercially available PMT was the side window 931 with electrostatic focussing only, designed by Zworykin and Rajchman (1939). Throughout the 1950s and 1960s, RCA extended the range of photocathodes through the research of A H Sommer (1980), who invented the bialkali and S20 photocathode types. During this period, Philips and EMI between them enjoyed a share of about 30% of the market, specializing, respectively, in fast and low-background PMTs. The development of negative affinity photocathodes in the 1970s made detection into the far infrared, beyond one micron, possible. Several specialist companies exploited these photocathodes mainly for military night vision purposes. Towards the end of the 1970s a Japanese company, Hamamatsu, began to dominate the side window market as RCA's position waned. The only manufacturer of side window PMTs is currently Hamamatsu, with a range of some 30 types of 10 mm and 30 mm geometries. The historical industry standards are still available, such as the 931, IP21, and the IP28; current product is similar in construction to the historical

4 Why photomultipliers?

versions but totally different in performance. The introduction of the S20 R928 in the 1970s signalled a significant advance with infrared sensitivity of two to three times that offered by the competitors, who to date have yet to find the key to manufacturing this outstanding side window PMT type. Hamamatsu produce both the world's smallest vacuum PMT, the μ PMT®, which is about the size of a thumbnail, and the biggest, which is a hemispherical tube with a diameter of 20"; they are the leaders of the PMT market with a share in excess of 90 %. Hamamatsu also have the leading position in the silicon detector market. RCA, Photonis, and Hamamatsu handbooks are available, all of which include a detailed historical sketch of PMT development (Engstrom 1980; Flyckt and Marmonier 2002; Hamamatsu 2007). Hamamatsu has set overall standards for consistency of performance, particularly with respect to gain and sensitivity, and nowadays it is unnecessary to allow a run-in period. Forty years ago the advice from experienced users was to switch on a week in advance of intended use. This has long been unacceptable as customers expect immediate stability of performance.

1.1.3 The statistical case for PMTs

A review article by Lallemand (1962), of the Paris Observatory, makes the case for PMTs by investigating the limitations imposed by well-known laws of physics. There are of course other reviews but Lallemand's treatment is novel, and his institute has the rare distinction of having made their own PMTs. His approach will be adopted, based initially on the analysis of a photocell, comprising a photocathode and an anode only; this is followed by the inclusion of a multiplier.

Consider the illumination of a photocathode with a random but steady source of light that produces a photocurrent, I_k . Random implies all frequency components are equally probable. By steady it is meant that the average current, measured over sufficiently long successive periods, is constant but superimposed on this current is a random component arising from the quantized nature of light and charge. This phenomenon is known as the Schottky effect. The rms noise current, $\langle i_s^2 \rangle^{1/2}$, is the shot noise and is given by

$$\langle i_s^2 \rangle^{1/2} = (2eI_k\Delta f)^{1/2}, \quad (1.1)$$

where e is the electronic charge, and Δf is the noise bandwidth. Note that shot noise only exists when a signal, I_k , is present: it may be interpreted as a random although quantifiable modulation, $\langle i_s^2 \rangle^{1/2}$, on I_k . Equation (1.1) is one of the more important relationships in the subject of light measurement, and its influence on detectivity is particularly felt whenever light levels are low. An inescapable consequence of (1.1) is that precise measurement of I_k presents a considerable challenge by the presence of this intrinsic noise, $\langle i_s^2 \rangle^{1/2}$. It is customary in PMT applications to perform current-to-voltage conversion at the anode by including a load resistor, R . However, all resistors are a source of EMF with associated noise, the magnitude of which is given by the Johnson noise formula:

$$\langle i_j^2 \rangle^{1/2} = \left(\frac{4kT\Delta f}{R} \right)^{1/2}, \quad (1.2)$$

where $k = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant, and T the absolute temperature, usually taken as 300 K. The thermal noise expression in terms of voltage follows from (1.2) as

$$\langle v_j^2 \rangle^{1/2} = (4kTR\Delta f)^{1/2}, \quad (1.3)$$

and it appears to predict an EMF with infinite noise for an open circuit (infinite resistance). However, the inescapable presence of stray capacitance, in addition to any deliberately added for bandwidth control, presents finite impedance to the signal. It is appropriate at this juncture to stress that, wherever noise is under consideration, the appropriate bandwidth is $\Delta f = 1/(4RC)$, and not the more familiar signal bandwidth, $1/(2\pi RC)$, although the numerical difference is small. Equation (1.3) then reduces to $\langle v_j^2 \rangle = kT/C$. It is useful to compare the relative contributions to noise from these two major sources, assuming typical values for the parameters involved.

Consider a photocurrent $I_k = 10^{-15}$ A, for example, which corresponds to an average flow of 6300 photoelectrons (pe)/s. This is a weak, but not ultra-weak signal, in PMT terms. Bandwidth, Δf , appears in both noise expressions, and for the present ratio metric analysis we take $\Delta f = 1$, $R = 1$, $T = 300$, and $\langle g \rangle = 10^6$ in equations (1.4)–(1.6), which follow. There is no loss in generality because the formulae readily scale. The mean gain, $\langle g \rangle$, of the multiplier, is enclosed in chevrons to emphasize its statistical nature. Shot noise originates from the current flow in the photocathode but it is enhanced by the multiplier gain, $\langle g \rangle$, when observed at the anode. On the other hand, Johnson noise is outside the influence of the PMT, since its source lies in an added load resistor. The contributions, taking Δf and R equal to unity, from shot and Johnson noise at the anode are

$$\langle i_s^2 \rangle^{1/2} = (2eI_k)^{1/2} \langle g \rangle = (2 \times 1.6 \times 10^{-19} \times 10^{-15})^{1/2} \times 10^6 = 18 \text{ pA}, \quad (1.4)$$

$$\langle i_j^2 \rangle^{1/2} = (4kT/R)^{1/2} = (4 \times 1.38 \times 10^{-23} \times 300)^{1/2} = 129 \text{ pA}. \quad (1.5)$$

The combined noise is given by adding contributions in quadrature, noting that (1.4) is signal dependent and (1.5) is not,

$$\langle i_{(s+j)}^2 \rangle^{1/2} = (\langle i_s^2 \rangle^2 + \langle i_j^2 \rangle^2)^{1/2} = (18^2 \times 129^2)^{1/2} = 130 \text{ pA}. \quad (1.6)$$

Equations (1.1), (1.2), and (1.6) are plotted in Fig. 1.1 as a function of R by adopting unit bandwidth, as is standard practice. Note how the procedure of adding in quadrature makes the net effect of the shot contribution negligible, in this case.

The two contributions are equal when

$$\frac{\langle i_s^2 \rangle^{1/2}}{\langle i_j^2 \rangle^{1/2}} = \left(\frac{2eI_k R}{4kT} \right) \times \langle g \rangle = 1. \quad (1.7)$$

6 Why photomultipliers?

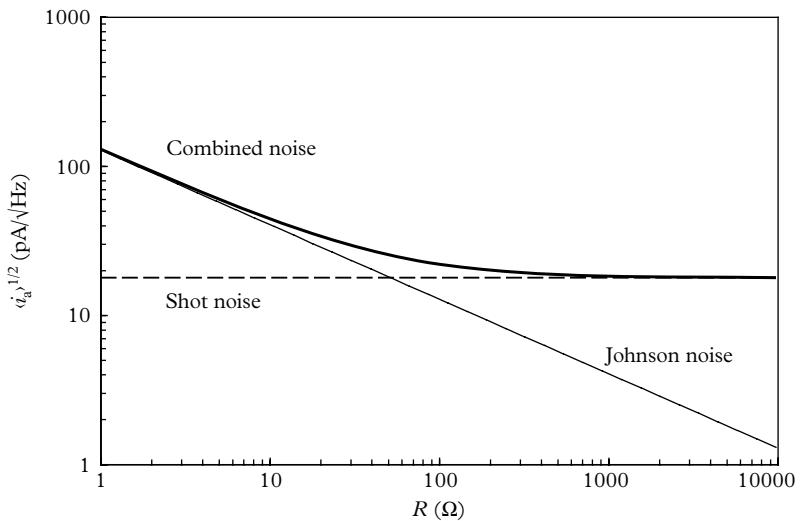


Fig. 1.1. Comparison of the magnitudes of shot and Johnson noise, at the anode, for an assumed cathode current of 10^{-15} A, and a multiplier gain of 10^6 . Note that the two noise sources are the same in magnitude for a load resistance of 50Ω —a value favoured for signal transmission via matched coaxial cable. The combined noise is the sum of the two sources taken in quadrature.

It follows from (1.7) for $I_k = 10^{-15}$ A, and $g = 10^6$, that R is 50Ω , and independent of Δf . The smallest measurable cathode current is determined by the Johnson noise in the load resistor. *The principal attribute of an electron multiplier is that it gives current amplification without a resistor.*

Handling fast, nanosecond rise time, output pulses from a detector demands a bandwidth in the region of 100 MHz and preferably a load resistor of 50Ω , for the reasons previously given. The combined noise is

$$\begin{aligned} \langle i_{(s+j)}^2 \rangle^{1/2} &= \left(\frac{4kT}{R} + 2eI_k(g)^2 \right)^{1/2} \Delta f^{1/2} \\ &= 2.56 \times 10^{-11} \times \sqrt{10^8} = 0.256 \mu\text{A rms}. \end{aligned} \quad (1.8)$$

With, in this example, equal contributions derive from the two noise sources.

The effect of the wide bandwidth is to increase the noise from 25.6 pA rms, at unity bandwidth, to 0.256 μA rms at 100 MHz bandwidth.

The gain required, such that multiplier noise will exceed Johnson noise is given by

$$g \geq \left(\frac{4kT}{2eI_k R} \right)^{1/2} = \left(\frac{0.052}{I_k R} \right)^{1/2} = \left(\frac{0.052}{10^{-9}} \right)^{1/2}, g \geq 7000. \quad (1.9)$$

1.1.4 DC detection with a PMT

A traditional PMT comprises a photocathode and a current multiplier within the same evacuated glass envelope. Amplification is achieved by an m -stage electron multiplier, based on secondary emission. The key parameter is the mean stage gain, δ , which results in an overall gain of δ^m at the output—referred to as the collector or anode. As we are considering an ideal case, all stages are assumed to have the same gain and to obey Poisson statistics. Gain may be expressed as either $\langle g \rangle$ or more specifically as δ^m , depending on the context. The terminology implies that the current amplifier acts on each photoelectron released from the photocathode by applying an overall multiplication factor of typically 10^6 . Thus, in this example, a charge of 1.6×10^{-19} C is increased to 0.16 pC by multiplication.

We now examine the evolution of noise as the signal passes each dynode stage. For a noiseless multiplier, the overall mean square noise after m stages would be the same as that at the photocathode, but enhanced by the overall gain, that is $\langle q_m^2 \rangle^{1/2} = (2eI_k\Delta f)^{1/2} \delta^m$. Since dynode multiplication is a random phenomenon, assumed to obey Poisson statistics, we must modify the shot noise formula to allow for this additional noise source, originating in the multiplier. The derivation is given in (4.11), requiring the inclusion of a factor $\delta^m(\delta^{m+1} - 1)/(\delta - 1)$, arriving at

$$\langle q_m^2 \rangle = 2eI_k\Delta f \left[\delta^m \left(\frac{\delta^{m+1} - 1}{\delta - 1} \right) \right]. \quad (1.10)$$

The signal-to-noise ratio at the output, $(S/N)_{op}$, divided by that at the input, $(S/N)_{ip}$, follows from (1.1) and (1.10) and is given by

$$\frac{(S/N)_{op}^2}{(S/N)_{ip}^2} = \frac{\delta^m(\delta - 1)}{\delta^{m+1} - 1}. \quad (1.11)$$

In practice, δ^{m+1} , greatly exceeds unity; hence,

$$\frac{(S/N)_{op}^2}{(S/N)_{ip}^2} = \frac{\delta - 1}{\delta} = ENF, \quad (1.12)$$

where ENF stands for excess noise factor (ENF), which represents ideal performance for an electron multiplier conforming to Poisson statistics, although the multiplier noise in real PMTs always exceeds this. The true ENF, free from any assumptions concerning its theoretical form, can be calculated from a pulse height distribution. This is obtained by illuminating the photocathode with a random source of single photons, and is known as the single-electron response (SER). Calculating the variance of an experimental distribution leads to the ENF. Single electron gain measurements are discussed fully in 5.8, and electron noise factors in 4.11. ENFs are relevant to many other light detectors, principally in the way they degrade resolution, as noted in Table 1.1

Figure 1.2 reveals three important properties of an electron multiplier: the S/N ratio decreases as more stages are considered; the S/N ratio benefits by increasing

8 Why photomultipliers?

Table 1.1 Resolution capability of various light detectors assuming a preamplifier with an ENC of 1000 electrons; η is the QE, and N is the number of photons in the signal.

Detector	η (%)	ENF	$\langle g \rangle$	Resolution
PMT	30	1.2	10^6	$(4/N)^{1/2}$
Fine-mesh PMT	30	2.0	10^6	$(6.7/N)^{1/2}$
MCPPMT	20	1.5	10^6	$(7.5/N)^{1/2}$
PIN diode	80	1	1	$[1.25/N + (1250/N)^2]^{1/2}$
APD	80	10	100	$[12.5/N + (12.5/N)^2]^{1/2}$
SiPM	30	1	10^6	$(3.3/N)^{1/2}$
VLPC	80	1	6×10^4	$(1.25/N)^{1/2}$
HPD	30	1	10^3	$[3.3/N + (3.3/N)^2]^{1/2}$
HAPD	30	1	10^5	$(3.3/N)^{1/2}$
Poisson limit				$(1/N)^{1/2}$

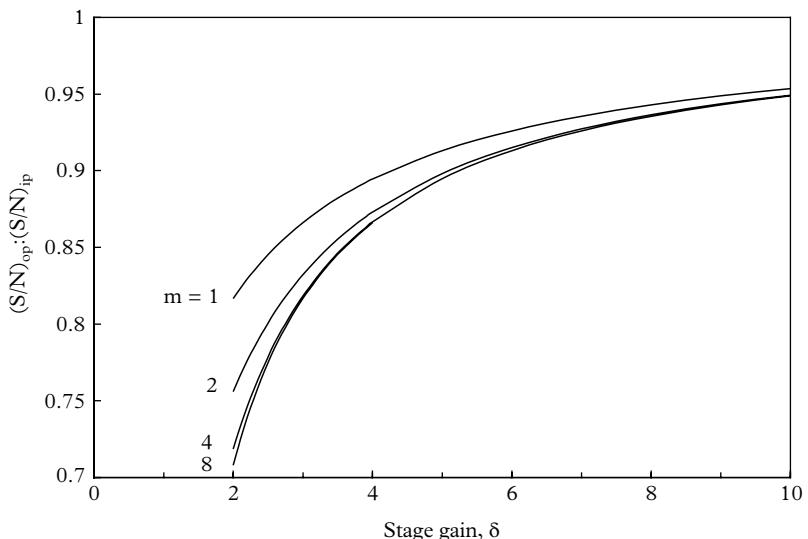


Fig. 1.2. S/N ratios, as a function of the number of dynodes for various stage gains, from (1.11). The curve for $m = 8$ is practically indistinguishable from that for $[(\delta - 1)/\delta]^{1/2}$. Note the S/N ratios rapidly tend to unity with increasing δ .

the stage gain, and the first stage, in particular, should be of high gain (see 4.10.3). The latter requirement is easily fulfilled by setting V_{k-d1} higher than the interdynode voltages—typically by a factor of 2 to 3. From (1.3) we note that the rms Johnson noise $\langle v_j^2 \rangle^{1/2}$ increases as $R^{1/2}$, but high multiplier gain permits the use of a relatively small resistor to overcome the contribution from this noise. Also, the choice of a low-value load resistor, R , is mandatory if wide bandwidth, and hence good high-frequency response, is required. Clearly, stray capacitance associated with the anode should be minimal to preserve bandwidth. Contributions to capacitance arise from internal wiring, the PMT socket, a circuit board if included, and unmatched signal transmission leads.

1.1.5 Detection of single photoelectrons

The discussion so far has centred on the measurement of currents but, if the signals under observation are feeble, and consequently photon detections are sufficiently separated in time, it is feasible to count individual photoelectron-initiated events, given sufficient multiplier gain. For circuit-analysis purposes, single photoelectrons may be treated as delta functions: the shape of the signal after multiplication depends on the time dispersion introduced by the multiplier, and on the bandwidth of any circuitry associated with the anode. Time dispersion can be measured, while the bandwidth is set by the user. PMTs with linear focussed dynodes have a bandwidth approaching 1 GHz, while for microchannel plate (MCP) devices Δf may reach 10 GHz. Noise calculations generally assume self-capacitance, C , associated with the anode of 10 pF, and a noise bandwidth of $\Delta f = 1/(4RC)$. From (1.2), the mean square of the Johnson noise at the anode load is

$$\langle v_j^2 \rangle = \langle i_j^2 \rangle R^2 = \frac{4kTR}{4RC} = \frac{kT}{C}. \quad (1.13)$$

A single photoelectron creates a signal on the anode capacitance of $v_a \sim e\langle g \rangle/C$. For the signal to exceed the Johnson noise,

$$\begin{aligned} \frac{e\langle g \rangle}{C} &\geq \left(\frac{kT}{C} \right)^{1/2}, \\ \langle g \rangle^2 &\geq \left(\frac{C}{e} \right) \left(\frac{kT}{e} \right). \end{aligned} \quad (1.14)$$

Note that $kT/e = 0.025$ V at room temperature, and C/e has the dimensions of volts^{-1} . From (1.14), the output signal has a magnitude equal to that of the Johnson noise at $\langle g \rangle = 1250$ for $C = 10$ pF; $\langle g \rangle$ would need to be an order of magnitude bigger in practice to clearly separate signal from background. Even at ten times the gain, the peak signal is only $e\langle g \rangle/C = 0.2$ mV—still uncomfortably small, even for the best, low-noise electronics. It is customary to operate at a gain in

10 Why photomultipliers?

excess of 10^6 to distance the signal at the anode from the effects of electromagnetic interference and power supply ripple. In so doing, the signal is still further removed from the influence of Johnson noise. It will already be appreciated that this treatment is approximate in two respects: first, it has been assumed that the output signal, v_a , is determined by φ and C only, with the influence of R ignored; second, there is noise associated with φ but the correction, according to (1.12) and Fig. 1.2, it is very small.

The foregoing treatment is intended to highlight the unique feature of an electron multiplier in separating signal from noise. This is achieved through resistorless current gain, from an ideal current generator. It is always important to bear in mind that PMTs do not generate ‘voltage signals’. These are created by adding a load resistor to the anode to perform current-to-voltage conversion. It is shown in 14.9.1 and 14.11.5 that one of the preferred circuits for use with PMTs is a transimpedance amplifier: the input impedance is close to zero, and direct connection to the anode is possible, by relocating the anode resistor to the feedback path of the amplifier.

There is an important difference between the amplification process of a PMT and that of an operational amplifier (OP-AMP), for example. Feedback in an electronics amplifier provides a means for selecting and stabilizing the gain, but this technique does not apply to PMT signals. Gain stabilization in PMTs can be achieved by other means for pulsed signals. An LED signal is arranged to illuminate the photocathode, without interfering with the primary signals. Ideally, the monitor signal is set outside the spectral range of interest. The technique employed in oil well logging with a NaI(Tl) scintillator is based on nullifying the shift in a ^{241}Am peak (59 keV), caused by the extreme range in operating temperature of up to 200°C, although a practical problem with LEDs is their temperature dependence.

1.1.6 Detection of multi-photoelectron signals

There are two modes of operation for light signals containing pulses of photons. A primary requirement is to determine the total charge produced at the anode by each event. This involves signal integration, which may be performed using passive components only or by using a charge-sensitive preamplifier. If the gain, collection efficiency, and the QE are known, it is possible to relate the charge in the anode signal to the incident pulse of photons. However, it is usually more desirable to attach a scale relevant to the application, rather than to the PMT. Hence, the accepted size unit in scintillation spectroscopy, for example, is MeV.

The relevant information in the second mode of operation is contained in the time signature of the event, making it important to preserve the transient shape of the light signal. This requires a relatively low-value resistor for the anode load, such as $50\ \Omega$, to mitigate the effects of stray capacitance. For example, with 10 pF, the RC time constant is 0.5 ns, or equivalently a signal bandwidth of 318 MHz (see 14.6.1). Consequently, transients with rise and fall times exceeding 1 ns suffer

only minimal degradation through bandwidth limitation. The choice of $50\ \Omega$ has the further benefit of matching, when signals are transmitted via coaxial cable, as previously mentioned. A rough estimation of the magnitude of the output signal may be derived from the following considerations. Assuming a gain of 10^6 and a load of $50\ \Omega$, a pulse of 2 ns width will generate a signal of $v_a \sim 50 \times 1.6 \times 10^{-19} \times 10^6 / (2 \times 10^{-9}) = 4\text{ mV/pe}$ —a signal that is measurable on any reasonable quality oscilloscope. The effects of stray capacitance can be further reduced by locating the load resistor in the feedback path of a transimpedance amplifier, ideally mounted on the base of a PMT. Applications in which preserving pulse shape is important are found in fluorescence decay, LIDAR, and pulse shape discrimination, for example.

1.1.7 Summary of PMT key attributes

- The smallest measurable light flux is limited by Johnson noise generated in an external resistor. According to (1.3) the rms noise scales as \sqrt{R} . The principal motivation for developing the electron multiplier is to provide current amplification without the use of a resistor. In practice, Johnson noise is seldom a concern.
- A multiplier can never improve the signal-to-noise ratio of incident signals. That is, increasing multiplier gain is not statistically equivalent to increasing the light flux.
- The multiplier has a diminishing effect on the signal-to-noise ratio as δ , the stage gain, is increased.
- Increasing gain at the first dynode is particularly beneficial.
- The signal-to-noise ratio improves with the adoption of fewer multiplier stages.
- It is feasible to detect and thus count single photoelectrons with a multiplier gain of the order of 10^6 . The photon-counting method discussed in 7.2.1 has many advantages over other techniques: one of these is the relative insensitivity to multiplier noise. With a sufficiently low detection threshold it is feasible to count the majority of output signals, particularly where the ENF is close to unity, as described in (1.12).
- Pulses generated by multi-photon events may be encoded in terms of total charge, or an associated unit, through calibration.
- The time signature of a transient light signal can be faithfully reproduced by exploiting the wide bandwidth characteristic of PMTs, with a gain-bandwidth product far exceeding those of the fastest OP-AMPS. Unless other considerations, such as non-linearity of response, fatigue, or microphonics apply, it is best to derive gain from the multiplier rather than electronically.

1.2 Other light detectors

1.2.1 Silicon PIN diodes

Silicon diodes are compact detectors intended primarily for high-light-level applications. There are over 100 different types, ranging from 0.1 to 20 mm in diameter. Self capacitance of large-area photodiodes restricts applications because of their low bandwidth. Capacitance can be reduced, however, by operating under reverse bias. These are unity gain devices with concomitant high noise and excessive dark current, although the latter reduces on cooling. Detection of gammas with inorganic scintillators, such as CsI and bismuth germanate BGO scintillators, is feasible at energies above about 100 keV.

1.2.2 APDs

The essential difference between PIN diodes and APDs is one of gain. PIN diodes are unity gain devices, while APDs are available with gain from 1 to 10^6 , depending mainly on the APD type. In PIN diodes the detected photon to electron–hole pair ratio is unity; in APDs photoelectrons excite further electrons, initiating an avalanche and, consequently, electron gain. Response time is faster than PIN diodes (e.g. 4 GHz for a Hamamatsu G8391-04 small area device). As for PIN diodes the capacitance and hence the noise can be reduced by applying reverse bias. Bandwidth is limited by self-capacitance, especially for large-area diodes. The gain–voltage relationship is not only temperature sensitive but is extremely steep in the preferred region of operation, thus demanding a highly stable voltage supply. Gain has an associated ENF ranging from 2 to 10, compared with ~ 1.2 for a typical vacuum PMT. Dark current is generally much lower in PMTs than in APDs, and these two parameters work in favour of the performance of a PMT. A range of over 100 different types from 0.04 to 40 mm diameter, offered by many manufacturers, is available in arrays and bespoke geometries. There is a selection of materials for fabricating semiconductor detectors, with devices covering the range 200 to 1700 nm, shown in Fig. 1.3.

EG&G was the first company to offer a high-gain photon-counting module, operating in charge saturation mode. The SPCM-AQ series of cooled photon-counting modules included detection diameters of up to 0.5 mm, with photosensitivity over the standard range of 400–1060 nm. Significant advances have been made in large-area APDs, particularly in the development by Hamamatsu of blue-sensitive and rad-hard diodes specifically for the Compact Muon Solenoid (CMS) experiment (Deiters *et al.* 2000). Large-area diodes, with a QE of 77 % at 400 nm, give performance with NaI(Tl) previously only attainable with PMTs. Measurements by Moszyński *et al.* (2000) are listed in Table 1.2 together with results for Radiation Monitoring Devices Inc. due to Farrell *et al.* (2000).

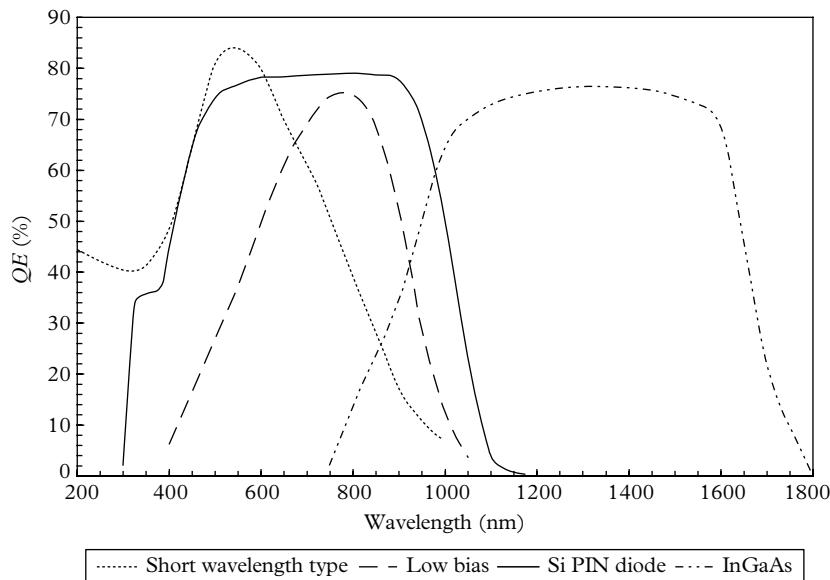


Fig. 1.3. QE versus wavelength for four different types of APD. It is customary to quote the sensitivity in millamps per watt.

Table 1.2 Resolution obtained with inorganic crystals and large-area APDs from Advanced Photonix (AP) and Radiation Monitoring Devices Inc. (RMD).

Type; diameter (mm)	Crystal (mm ²)	E (keV)	Res (%)	NE (keV)
AP, 630-70-73-500; 16	CsI(Tl) 9 × 9	662	4.8	10
AP, 630-70-73-500; 10	NaI(Tl) 10 × 10	122	8.4	6
AP, 630-70-73-500; 10	NaI(Tl) 10 × 10	59	11.3	6
RMD; 40	CsI(Tl) 25 × 25	662	13.5	170
RMD; 14	CsI(Tl) 13 × 13	662	7.0	60
RMD; 9	CsI(Tl) 8 × 8	662	6.0	45

Noise edge (NE), listed in the last column of Table 1.2, is a specification parameter for classifying high-temperature operation with PMTs, and equally valid for present considerations. It is a measure of the encroaching tail of dark counts and noise into the signal spectrum of pulse heights (discussed in 12.2.2). Gamma energy peak heights need to lie significantly distanced from the NE to be resolved.

1.2.3 SiPMs

There are currently more than ten manufacturers of SiPMs: principally, Hamamatsu, Photonique, STM, SensL, RMD, and Perkin Elmer. As stated in the excellent review article by Musienko (2009): ‘Every producer has invented a name for their device, amongst which we have MRS, MRSAPD, MAPD, SiPM, SSPM, SPM, G-APD’. The first single pixel G-Mode APD, manufactured by EG&G, is the one described by Haitz *et al.* (1963). Micropixel APDs (MTS APD) were patented by Golovin *et al.* (1989), and the first commercial devices followed a few years later in Russia. The SiPM is a multi-pixel APD with all channels terminating on a common substrate. A 2-D array of pixels, as small as $10\text{ }\mu\text{m}$, constitutes a packaged detector of typically $1 \times 1\text{ mm}^2$. The number of pixels per unit area ranges from 100 to 100,000. Each pixel operates at a bias voltage of about 50 V, corresponding to $\sim 15\%$ above the breakdown voltage in the so-called Geiger mode. A quenching resistor is included in each pixel to terminate the discharge. Pixels generate signals with common amplitude, because every avalanche runs into saturation. Hence, a multi-photon flash produces a comb-shaped response by separating individual photon detections, depicted in Figs 1.4 and 1.5. In excess of 10 pe can be resolved and hence counted—a consequence of an ENF close to unity.

Sensitivity is usually quoted in terms of photon detection efficiency (PDE), which is the product of QE, fill factor (dead space), and avalanche probability.

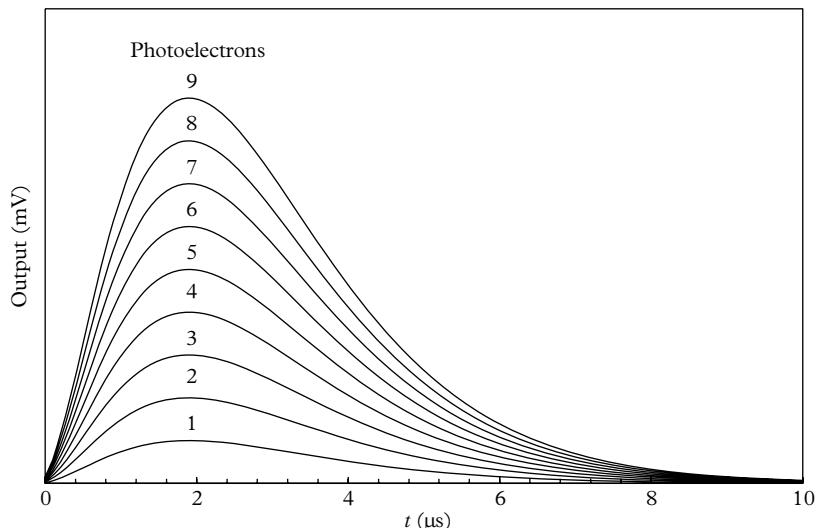


Fig. 1.4. A representation of the output after shaping and filtering, as seen on an oscilloscope. In reality there are signals in the valleys between the traces, but their number is small.

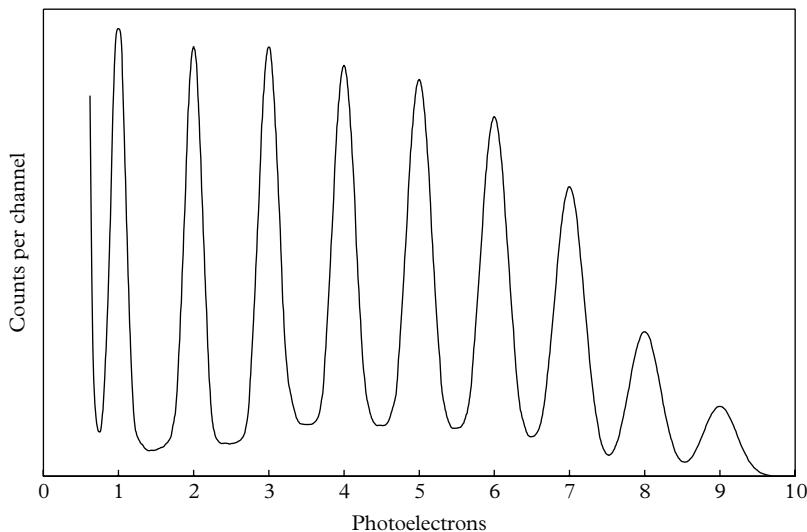


Fig. 1.5. A multi-photoelectron pulse height distribution for which the mean number of photoelectrons per flash is 4. The steeply falling edge on the left of the distribution is referred to as the pedestal or noise edge.

The realizable QE is about half that of a single-channel APD. The original SiPMs exhibited high red sensitivity with a spectral response similar to the single-channel APDs. An important development was the enhanced blue-green response introduced to match the light output from scintillators; PDE curves are shown in Fig. 1.6 for N-type Hamamatsu devices and the P-type detector from CPTA/Photonique. Note the N-type has similar sensitivity to a vacuum PMT in the blue-green region of the spectrum but with a superior response to red light.

Linearity with regard to light input improves by increasing the number of pixels, and with uniformity of illumination. The signal output of an individual pixel is fixed, regardless of the number of coincident photons received; hence the improvement in dynamic range which follows from uniform illumination. Linearity, for the Hamamatsu C12661 modular series, extends to 10^7 single-photon counts/s, for 15 μm pixels (see e.g. Hamamatsu n.d., p. 38).

Dark count rates are high in these devices, especially at room temperature: 100 kHz to several megahertz per square millimetre is not uncommon. They reduce exponentially by a factor of 2 per 8 °C drop in temperature. The time distribution for dark counts shows an excess of undersized intervals below 1 μs because of correlations. Afterpulse rates in the Hamamatsu S12571-050C were initially high but modified designs reduced this contribution to 1 to 2%, provided that overvoltage is kept below 3 V. Dark count rates are strongly dependent on threshold, with a decrease of four orders at a threshold of 2.5 photoelectrons equivalent.

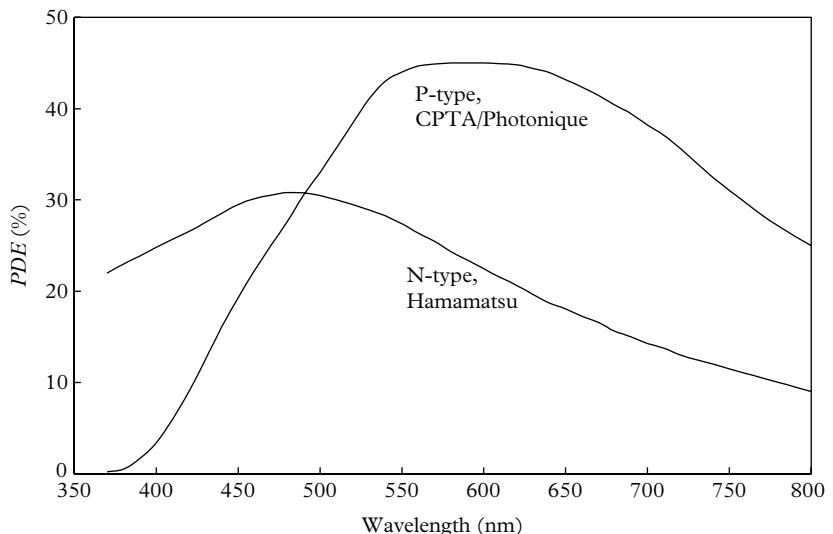


Fig. 1.6. PDE is similar in concept to QE as defined for a PMT, but includes collection efficiency; P-type detectors are best suited to applications involving red and infrared light, while the N-type is preferred for scintillator applications.

The excess noise factor, $ENF = 1 + \text{var}(g)/\langle g \rangle^2$, deduced from a single photoelectron distribution such as the first peak in Fig. 1.5, can be as low as 1.03 at a gain of 10^6 . It is this property that imbues an SiPM with its multiple-photoelectron-resolution capability. Similarly, crosstalk between channels has seen a significant decrease to about 1% for 2 V of overvoltage. Time resolution with single photons of 100 ps full width at half maximum (fwhm) has been reported (Musienko 2009). Detection requirements for small PET scanners based on a lutetium oxyorthosilicate (LSO) scintillator, and wavelength-shifting fibres imbedded in tiles (see 3.12) makes the small, pixelated SiPMs ideal for these applications. Recovery time of an activated pixel has been studied by Britvitch and Renker (2006), who measured a selection of devices from different manufacturers; they found that recovery ranged from 0.1 to 600 μ s with a strong dependence on temperature and bias voltage. The effective QE is halved, because of dead space between pixels known as the fill factor. Nevertheless, sensitivity exceeds that of a PMT, especially for infrared light.

1.2.4 Summary of silicon key attributes

The important characteristics common to all the silicon based detectors are:

- availability of arrays and bespoke geometries
- high QE, approaching 80 % within the range 400 to 900 nm

- robust and resistant to damage from extreme light levels
- wide dynamic range
- insensitive to magnetic fields
- rad-hard options available
- can be operated unbiased
- capacitance can be reduced by a factor of 4 by reverse biasing

1.2.5 Visible-light photon counters

Visible-light photon counters (VLPCs) are solid state detectors that have operating characteristics similar to those of SiPMs, although the technology is different. QE is high, and gain of 40,000, as described by Bross *et al.* (2002) for the DØ detector at Fermilab, can be achieved. Used with scintillating fibres (shifted to orange wavelengths) in the presence of a high rate of background counts of 100 kHz, these detectors need to be cooled to within a few degrees of absolute.

1.2.6 CCDs

CCDs find application in a wide range of industrial (electron microscopes), astronomical (satellite and ground based) and scientific (scintillating fibres) fields. CCDs are signal integrating devices providing excellent image quality. Cooling leads to detection down to tens of photons.

1.2.7 Hybrid photodetectors and APDs

These PMTs are described as hybrid because they combine two technologies: a solid state multiplier and a semi-transparent photocathode within the same evacuated envelope. This configuration is attractive because it combines the large detection area of a classical PMT with the energy and spatial resolution of a segmented silicon device. The solid state multiplier is either a large-area silicon PIN diode, in which case the product is a hybrid photodetector (HPD), or, if an APD is incorporated, the device is referred to as a hybrid APD (HAPD). The diode of whichever configuration is operated up to -20 kV with respect to the photocathode. Each photoelectron produces one electron-hole pair per 3.6 eV energy loss in silicon, generating an electronic pulse charge of about $2778 \times 1.6 \times 10^{-19}$ C for a 10 kV energy loss. From charge-sensitive preamplifier considerations (see 14.9.1), it is clear that an amplifier with 300-electron noise would not noticeably affect resolution. Multi-photoelectron resolution is poorer than that of the SiPM, and yet still impressive, with a resolution of up to at least ten peaks. The silicon anode can be segmented, for example, into 50 μm pixels for imaging purposes. The preferred size range for HPDs is from 2 to 13" in diameter, including hemispherical PMTs. The HAPD developed for the MAGIC-II

atmospheric Cerenkov telescope had a 1" GaAsP photocathode with a wavelength shifter providing a QE of >30 % over the range 250 to 700 nm. These devices were commercially first available from DEP, but the CERN PH department have, in the past, manufactured their own. Currently, the only manufacturer is Hamamatsu. The advantage these detectors have over traditional PMTs is their relative insensitivity to magnetic fields achieved as a result of an accelerating voltage in excess of 20 kV, compared with 300 V, V_{k-d1} , for a PMT. The ENF is similar to that of a PMT, the response is fast, and the target silicon diode may be pixelated. Production costs are higher than those for PMTs.

1.3 Pulse height resolution

An important property of a light detector is its ability to separate transient events that contain approximately the same mean number of photons. The figure of merit is based on the relative width of the pulse height distribution derived from a steady pulsed light source, such as an LED. If the mean number of photons per pulse is N then the standard deviation, σ_N , of the distribution, is \sqrt{N} , and the relative standard deviation, the resolution, is $\sigma_N/N = 1/\sqrt{N}$ on assuming Poisson statistics. The resolution measured at the output of a detector only approaches $1/\sqrt{N}$ since some photons fail to produce photoelectrons and, where gain is concerned, there is a further loss in resolution. Detector resolution is determined by the number of photoelectrons produced, given by $N_{pe} = N\eta\varepsilon F$, where η is the QE at the wavelength of interest, ε is the fill factor, which accounts for inter-pixel dead space in imaging devices, and F is the collection efficiency for photoelectrons. Note that η is always less than 100%, whatever detector is under consideration; ε applies to certain pixilated photocathodes intended for imaging purposes. Collection efficiency, F , is relevant to vacuum photon detectors such as PMTs, HPDs, HAPDs, and MCPPMTs. All detectors reflect a portion of incident light, to the extent of 10 to 30%, depending on wavelength, polarization, and incident angle. These three considerations are only partially included in η , as quoted by manufacturers. Except for PIN diodes, all light detectors considered here provide gain, but the generating mechanism is statistical in nature and hence noisy. For an ideal PMT, with all stage gains equal to δ , $ENF = \delta/(\delta - 1) = 1.11$ for $\delta = 10$, for example. In reality the ENF is ~ 1.2 for a quality PMT (see Figs 5.5 to 5.7) and this value will be assumed for present purposes. Lastly, most detector systems include a shaping preamplifier, specified by an equivalent noise charge ENC , quoted in electrons. The pulse height resolution, adding the two noise sources in quadrature, is

$$\frac{\sigma_N}{N} = \left[\frac{ENF}{N_{pe}} + \left(\frac{ENC}{N_{pe}(g)} \right)^2 \right]^{1/2} \quad (1.15)$$

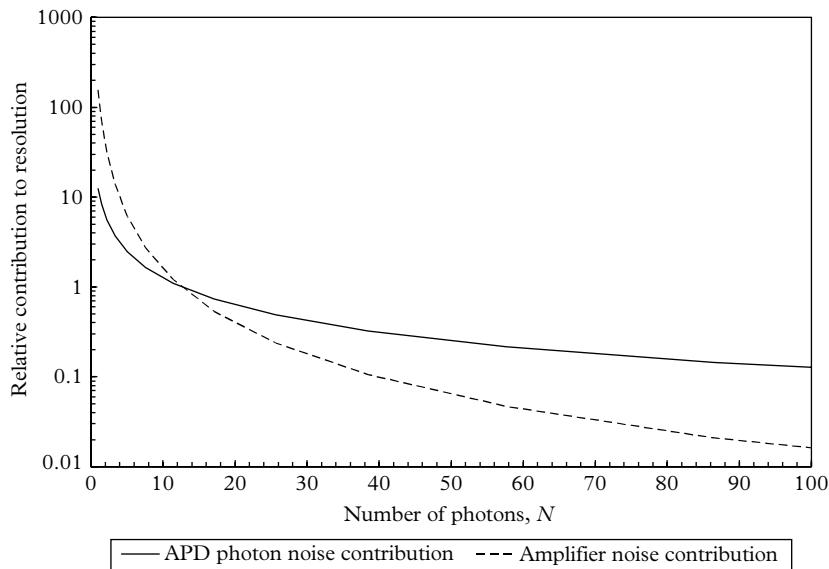


Fig. 1.7. Contributions to overall resolution determined by separately considering the two noise sources in the pulse height resolution formula (1.15). Based on $ENF = 10$, amplifier noise = 1000 electrons, $\langle g \rangle = 100$, and $N_{pe} = 0.8 N$.

Of the two terms within the square brackets, there is always a contribution from the first, whereas the second is relatively unimportant for detectors with high gain. As an example, consider the relative contributions of each term for $N = 20$ photons, and the parameters quoted in Fig. 1.7 for an APD:

$$\begin{aligned}\frac{\sigma_N}{N} &= \left[\frac{10}{20 \times 0.8} + \left(\frac{1000}{0.8 \times 20 \times 100} \right)^2 \right]^{1/2} \\ &= (0.625 + 0.391)^{1/2}.\end{aligned}$$

Individual contributions from the two terms are illustrated in Fig. 1.7; the benefit of this exercise lies in being able to specify the maximum acceptable noise for any preamplifier under consideration. This example shows that detection of a 20-photon pulse is feasible with the selection of an amplifier with a noise of 1000 electrons. The concept of quoting amplifier noise in terms of electrons is explained in 14.10.2. Table 1.1 compares the resolution capability of a range of detectors, following the method adopted by Arisaka (2000). The second term in the resolution formula (1.15) is omitted in the tabular results whenever its contribution is negligible for all N . This is certainly true for a PMT for which the resolution is given by

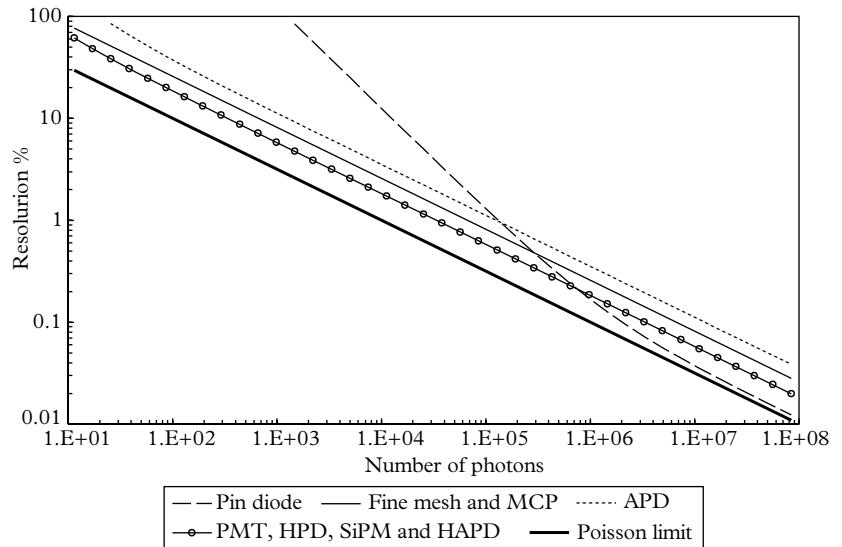


Fig. 1.8. Resolution for various detectors as a function of the number of photons per pulse from Table 1.1.

$$\frac{\sigma_N}{N} = \left(\frac{ENF}{N_{pe}} \right)^{1/2} = \left(\frac{1.2}{0.3 \times N} \right)^{1/2} = \left(\frac{4}{N} \right)^{1/2}. \quad (1.16)$$

Resolution formulae for fine-mesh PMTs and MCPPMTs follow in a similar manner. Equation (1.17) shows that amplifier noise is a major contributor to a PIN diode, of unity gain, since

$$\frac{\sigma_N}{N} = \left(\frac{1}{0.8 \times N} + \left(\frac{1000}{0.8N \times 1} \right)^2 \right)^{1/2} = \left[\frac{1.25}{N} + \left(\frac{1250}{N} \right)^2 \right]^{1/2}. \quad (1.17)$$

The remaining entries in Table 1.1 follow from similar reasoning. The relationships in the last column are plotted in Fig. 1.8 after normalization. Note that the relative positioning of the curves depends on the assumed parameters—especially η , for which the maximum values for each particular detector have been assumed. The entries highlight the importance of high gain, even if noisy, as for an APD. Note that at sufficiently high light levels, a PIN diode will provide the best resolution of all detectors if used in combination with a low-noise preamplifier. The assumed noise figure of 1000 electrons describes an uncooled amplifier of modest performance; cooled devices offered by AMPTEK, for example, have a quoted noise of 20 electrons.

It is convenient in scintillation spectroscopy, and in other disciplines, to define resolution in terms of the relative fwhm of a pulse height distribution, thereby

avoiding arithmetic computation. This gives rise to a factor of $2 \times (2 \ln 2)^{1/2}$ $\sigma = 2.35\sigma$ for a normal distribution, with standard deviation, σ , for converting from one resolution specification to the other (further details of this conversion can be found in 8.4.2). The important consideration is consistency: for present purposes we take resolution as σ_N/N but it is easily expressed in terms of fwhm when required. It is also useful to disentangle the two contributions in a pulse height resolution formula, particularly for an APD. The intersection of such curves gives an indication of the minimum noise performance required of any proposed preamplifier.

1.4 Position resolution

The detectors listed in Table 1.1 are capable of imaging, with the exception of the traditional vacuum PMT. This attribute can only be realized in devices that are sensitive to the point of detection on the photocathode. The design of the multiplier and the anode must ensure the transfer of this position to the anode (or anodes), with the minimum distortion. The MCPPMT, one of the earliest imaging detectors, is based on proximity focussing. That is, the window and photocathode are everywhere parallel to the MCP and disposed close to it. A kilovolt biasing voltage preserves the (x, y) coordinates of each emitted photoelectron and provides gain for the MCP.

Fine-mesh and metal-channel electron multipliers, included in proximity-focussed PMTs, give positional resolution down to the sub-millimetre level. The flat panel, 12-stage H8500 series from Hamamatsu with an overall length of 33 mm, has eight by eight separate anodes. Detailed information on position-sensitive PMTs may be found in the Hamamatsu PMT tube book (Hamamatsu 2007, Section 6.1).

1.5 Signal-to-background considerations

Detectivity and resolution are arguably the most important parameters where light detectors are concerned. The ability to detect and quantify weak signals, referred to as the detectivity, also depends on the level of dark counts or dark current. The discussion in 1.4 addressed the question of resolution by adopting a statistical approach embracing the quantized nature of light signals, including the effect of noisy gain and a noisy preamplifier—an attribute of all detectors except PIN diodes.

Certain detectors, particularly the semiconductor ones, suffer from high dark current, while vacuum devices such as PMTs may have dark currents some five orders lower, for the same detection area. Dark current and dark counts affect both

detectivity and resolution; although subtraction may lead to the signal component, statistical and other effects set a limit to precision. Detailed consideration of these topics is reserved for Chapter 6.

The analysis followed in Table 1.1 is recommended as a means for selecting the best detector for a given application; also, equally important, a signal minus background exercise should be performed. Where counts are measured, there is a well-known formula for the optimal division of time between making a signal plus background measurement and background alone determination. This is covered in detail in 7.4, with an example in Fig. 7.8 of recovering a signal of 0.041 counts/s from a background of 0.459 counts/s.

REFERENCES

- Arisaka, K. (2000). New trends in vacuum-based photon detectors. *Nucl. Instr. and Meth. in Phys. Res. A*, **442**, 80–90.
- Britvitch, I. and Renker, D. (2006). Measurements of the recovery time of Geiger-mode avalanche photodiodes. *Nucl. Instr. and Meth. in Phys. Res. A*, **567**, 260–3.
- Bross, A. *et al.* (2002). Characterisation and performance of visible light photon counters (VLPCs) for the upgraded DØ detector at the Fermilab Tevatron. *Nucl. Instr. and Meth. in Phys. Res. A*, **477**, 172–8.
- Deiters, K. *et al.* (2000). Properties of the most recent avalanche photodiodes for the CMS electromagnetic calorimeter. *Nucl. Instr. and Meth. in Phys. Res. A*, **442**, 193–7.
- Engstrom, R. W. (1980). *Photomultiplier Handbook*. RCA Corporation, Lancaster, Pa.
- Farrell, R. *et al.* (2000). APD arrays and large area APDs via a new planar process. *Nucl. Instr. and Meth. in Phys. Res. A*, **442**, 171–8.
- Flyckt, S.O. and Marmonier, C. (2002). *Photomultiplier Tubes: Principles and Applications*. Photonis, Brive.
- Golovin, V. *et al.* (1989). Patent, #1702831 of Russia, priority from 1989.
- Haitz, R. *et al.* (1963). Avalanche effects in silicon *p–n* junctions. 1. Localized photomultiplication studies on microplasmas. *J Appl. Physics*, **34**, 158–91.
- Hamamatsu (n.d.) *MPPC®, MPPC Modules, Hamamatsu Technical Information*. Hamamatsu Photonics K. K., Hamamatsu City.
- Hamamatsu (2007). *Photomultiplier Tubes: Basics and Applications, Third Edition (Edition 3a)*. Hamamatsu Photonics K. K., Hamamatsu City.
- Lallemand, A. (1962). ‘Photomultipliers’. In *Astronomical Techniques*, Volume 2, Chapter 6, ed. W. A. Hiltner, Chicago University Press, Chicago, 126–56.
- Lubsandorzhiev, B. K. (2006). On the history of photomultiplier tube invention. *Nucl. Instr. and Meth. in Phys. Res. A*, **567**, 236–8.
- Moszyński, M. *et al.* (2000). Large area avalanche photodiodes in X-rays and scintillation detection. *Nucl. Instr. and Meth. in Phys. Res. A*, **442**, 230–7.
- Musienko, Y. (2009). Advances in multipixel Geiger-mode avalanche photodiodes (silicon photomultipliers). *Nucl. Instr. and Meth. in Phys. Res. A*, **598**, 213–6.

- Sommer, A. H. (1980). *Photoemissive Materials*. John Wiley and Sons, Inc., New York.
- Zworykin, V. K., Morton, G. A., and Malter, L. (1936). The secondary emission multiplier: A new electronic device. *Proc. IRE*, 24, 351–75.
- Zworykin, V. K. and Rajchman, J. A. (1939). The electrostatic electron multiplier. *Proc. IRE*, 27, 558–66.

2

Photocathodes

2.1 Introduction

The colour of an end window photocathode provides an immediate check on whether a PMT is alive or dead. If it is ‘down to air’, the inside of the window will appear colourless and the device is useless. The colour of the photocathode viewed under white light also reveals some of the basic physics of light detection. Bialkali photocathodes appear thin and yellow to the human eye when viewed under white light. That is, it is relatively easy to see the internal structure of a PMT under such conditions. The blue component of white light is preferentially absorbed by the photocathode, while the longer wavelengths are relatively unattenuated and hence visible. On the other hand, S20 photocathodes appear relatively thick and sometimes metallic when viewed in white light; this is indicative of high reflectivity at visible wavelengths. The quantity of light transmission through a photocathode is complementary to absorption, and consequently to low QE. High-quality photocathodes were traditionally manufactured based on colour changes, in addition to continuous monitoring of photocurrent. More reproducible photocathodes, generally of higher QE, are produced by automating the entire activation process, replacing human judgement by automatic parameter monitoring. How this is done has remained confidential to the companies that employ automated methods.

Research into the optical properties of photocathodes and their characterization is a key activity for manufacturers of photo-emissive devices. Such knowledge is also important to experimentalists involved in the more demanding applications of PMTs, for which light detection efficiency may ultimately decide the feasibility or otherwise of a proposed detector system. Astrophysics experiments, for example, may involve the use of thousands of PMTs immersed in water, ice, liquid argon, or air. The number of photons collected sets the minimum detectable energy, which in turn leads to a lower limit to the quantity of PMTs needed. Detection of wavelengths in the region of 900 nm has always been a challenge because of low QE and high dark current. The primary reason for the former is the transparency of the photocathode at long wavelengths, as previously mentioned. Most of the light passes through the photocathode and is undetected, unless some way can be found to recycle it. Ingenious methods have been developed since Rambo (1964) first achieved an enhancement in detectivity through optical means; this subject is fully discussed in 3.13.

The primary concern in this chapter is the optical characteristics of end window photocathodes. Comparatively little has been published on the topic and even less on side window photocathodes. The reasons for the latter relate mainly to difficulties in gaining optical access to a side window photocathode. Specifically, this type of photocathode lacks symmetry and is remote from the curved PMT entrance window, whereas an end window photocathode is

26 Photocathodes

fabricated directly on to a glass faceplate, usually flat, thin, and ideally positioned for receiving incident light. Although there remain difficulties in making optical measurements on this type of PMT, they are considerably less severe than those for side window types. Standard window materials are borosilicate glass, fused silica (quartz), and the less common, magnesium fluoride and sapphire, all with a refractive index in the region of 1.5. Every photocathode is associated with a glass window of the types previously referred to, and theoretical optical studies must allow for this. The basic topics of reflection and refraction are covered in 2.4, concerning the behaviour of light in mixed dielectric media.

2.1.1 Solid angles

Solid angles are involved in photometric quantities, such as luminous intensity (lumens per steradian), covered in 2.3.2. They are also relevant to light propagation in layered media of different refractive indices and to scintillators and light guides. These topics are treated in Chapter 3 and it is important to understand the concept of solid angle.

In two dimensions, illustrated in Fig. 2.1, an angle, in radians, is the length of the arc of a unit circle radius subtended by the angle. In three dimensions, a solid angle measured in steradians, is the area of the unit sphere subtended by the solid angle.

The partial area of the sphere subtended by θ is the solid angle, given by

$$\int_0^\theta 2\pi R^2 \sin\theta \, d\theta = 2\pi(1 - \cos\theta)$$

for a sphere of unit radius. The relative solid angle follows by dividing by the area of a unit sphere to give

$$\Omega = \frac{1}{2}(1 - \cos\theta). \quad (2.1)$$

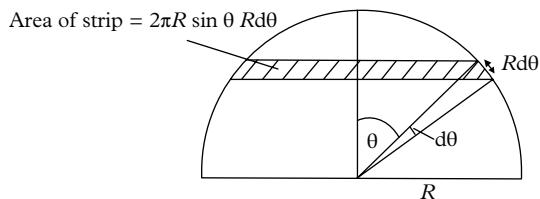


Fig. 2.1. The concept of solid angle, Ω .

2.2 Fundamentals of photosensitivity

2.2.1 The electromagnetic spectrum

The electromagnetic spectrum covers a range of 25 orders of magnitude in frequency, extending from high-energy gamma rays of cosmic origin to extra low-frequency radio waves. The sub-spectrum of interest in PMT applications covers less than a decade in wavelength. We now examine the relationships between various parameters, starting from the fundamental equation relating photon energy $E(\lambda)$ to the wavelength, λ . The energy carried by a photon is inversely proportional to its wavelength and the two quantities are related as follows:

$$\begin{aligned} E(\lambda) &= \frac{hc}{\lambda} = \frac{6.63 \times 10^{-34} \times 3 \times 10^8}{\lambda_m} \\ &= \frac{1.987 \times 10^{-25}}{\lambda_m} \text{ J}, \end{aligned} \quad (2.2)$$

where:

h is the Planck constant = $6.63 \times 10^{-34} \text{ J} \cdot \text{s}$

c is the speed of light = $3 \times 10^8 \text{ m/s}$

λ_m is the wavelength of the radiation, expressed in metres

From (2.2),

$$\text{a flux of 1 photon/s} = \frac{1.987 \times 10^{-16}}{\lambda_{\text{nm}}} \text{ W}, \quad (2.3)$$

where for convenience λ is now expressed in nm, λ_{nm} , and

$$1 \text{ pe/s} = 1.602 \times 10^{-19} \text{ A}. \quad (2.4)$$

The spectral responsivity, σ , of a photocathode is the photocurrent produced by 1 W of incident radiation, for a photocathode QE of $\eta(\lambda)$:

$$\sigma = \frac{1.602 \times 10^{-19} \eta(\lambda)}{\left(\frac{1.987 \times 10^{-16}}{\lambda_{\text{nm}}} \right)} \text{ A/W} = \frac{\lambda_{\text{nm}} \eta(\lambda)}{1.24} \text{ mA/W}. \quad (2.5)$$

For example, the QE of a typical bialkali photocathode is 25 % at 400 nm, which equates to

$$\sigma = \frac{400 \times 0.25}{1.24} = 80.6 \text{ mA/W}. \quad (2.6)$$

We see that the figure of merit of a photocathode can be expressed either as QE, or, as in (2.5), in millamps per watt. In general, QE is the appropriate parameter in photon counting, since it indicates the probability of producing a photoelectron per incident photon, and thus the expected number of pulses per second at the anode. QE is also relevant in scintillation spectroscopy, and other pulsed applications, as it relates to pulse height resolution. On the other hand, where steady or quasi-steady light sources are concerned, the appropriate parameter is millamps per watt, which is the photocurrent per incident watt of radiation. It is fitting to draw attention to the remarkable magnitude of 1 W of optical radiation: from (2.3) a HeNe laser emitting a power of 1 mW, for example, produces a flux of 3.2×10^{15} photons/s of wavelength 633 nm—well outside the operating range of a PMT.

2.2.2 Photoelectric phenomena

The essence of Einstein's work on the photoelectric effect is the equation for determining the maximum energy of an emitted photoelectron.

$$E = h\nu - (E_G + E_A) = \varphi_p. \quad (2.7)$$

The form of this equation is the same for metals and for semiconductors. Valence-band electrons can only absorb a quantum of radiation if the energy of the photon is at least equal to the band-gap energy, E_G . For photoemission to occur an additional energy, E_A , the electron affinity, must be provided; $(E_A + E_G)$ is the photoelectric work function of the material, φ_p , which is different in magnitude from the thermionic work function, φ_{th} (the two work functions are sometimes incorrectly used interchangeably). The minimum photon energy for photoemission can be predicted by expressing (2.2) in terms of electron volts (eV). Dividing by the electronic charge leads to

$$E(\lambda)_{\text{eV}} = \frac{1240}{\lambda_{\text{nm}}}. \quad (2.8)$$

We see at once that a wavelength of 1 μm , close to the long wavelength limit of detection by PMTs, corresponds to 1.240 eV; this energy is insufficient for producing photoelectrons from the majority of currently available photocathode types. Generation of 2 pe becomes energetically possible at short wavelengths but the response of the PMT is mitigated by the absorptance of the PMT window, rather than by the cathode sensitivity (see e.g. Fig. 2.17). For photon energies greater than $2(E_A + E_G) \sim 4$ eV, corresponding to a threshold of about 200 nm, pair production is energetically possible and some spectral response curves show evidence of this. Note that (2.8) does not predict the efficiency for producing electrons at any wavelength; this has to be determined by measurement. The detectable range of radiation for PMTs is very narrow, spanning approximately 100 to 1000 nm—barely an order of magnitude. Table 2.1 correlates spectral colours, wavelengths, and energy in electron volts.

Table 2.1 The range of wavelengths accessible to a PMT, and their colour designations. Equivalent energies are calculated from (2.8). The visible range extends from 390 to 770 nm.

Colour	Wavelength (nm)	Energy (eV)
Near infrared	770–1500	0.83–1.61
Red	622–770	1.61–2.00
Orange	597–622	2.00–2.08
Yellow	577–597	2.08–2.15
Green	492–577	2.15–2.53
Blue	455–492	2.53–2.73
Violet	390–455	2.73–3.19
Near ultraviolet	300–390	3.19–4.14
Far ultraviolet	100–300	4.14–12.4

2.2.3 Photoelectron energy distribution

Equation (2.8) indicates a long wavelength cut-off for energies below ($E_A + E_G$). For a Cs_3Sb (S11) photocathode, for example, this is about 2 eV, according to Sommer (1980), with a corresponding wavelength detection limit of about 600 nm. These photoelectrons barely escape the photocathode, whereas those produced by shorter wavelengths are emitted with an angular distribution and a spectrum of energies. The shape of photoelectron energy spectra is noticeably irregular, possibly reflecting the measurement difficulties rather than reality. The results of Spicer (1961), concerning the response of S20 photocathodes to blue and UV light, can be summarized as follows: the most probable emission energy is ~0.7 eV with a high-energy tail extending to 70% of the energy of the initiating photon. For example, a far-UV photon of energy 4.4 eV produces photoelectrons with a maximum energy of 3.1 eV. While photoelectron energy spectra have been published for commercially available photocathodes, there appears to be no information on angular distributions—also reflecting difficulties in making such measurements. The energy and angular distribution of photoelectrons contribute to the minimum spot size on the first dynode, with consequences for fast timing and collection efficiency, discussed in 8.2.1. Early versions of box-and-grid designs suffered particularly from defects caused by low extraction field gradients at the photocathode. These considerations must be included in any serious electron-optical simulation for which manufacturers generally choose a launch energy of ~1 eV, and assume a Lambertian angular distribution for such computations.

2.3 Spectral radiation and photometric units

2.3.1 Spectral radiant energy

In 2.2.1 we derived a basic formula relating photocurrent produced by radiation of a single wavelength, λ , incident on a photocathode of QE, $\eta(\lambda)$. We wish to broaden this to a radiation source producing a spectrum of wavelengths, $W(\lambda)$, incident on a photocathode with a spectral response $\eta(\lambda)$. The average power radiating from a light source is

$$P = P_0 \int_0^\infty W(\lambda) d\lambda, \quad (2.9)$$

where P_0 is the power in watts per unit wavelength at the peak of $W(\lambda)$, normalized to unity. The resulting photocathode current, I_k , is

$$I_k = \sigma P_0 \int_0^\infty W(\lambda) \eta(\lambda) d\lambda. \quad (2.10)$$

From (2.9),

$$I_k = \frac{\sigma P \int_0^\infty W(\lambda) \eta(\lambda) d\lambda}{\int_0^\infty W(\lambda) d\lambda}. \quad (2.11)$$

If the input light is a filtered version of $W(\lambda)$, according to some function $f(\lambda)$, (2.11) becomes

$$I_k = \frac{\sigma P \int_0^\infty W(\lambda) \eta(\lambda) f(\lambda) d\lambda}{\int_0^\infty W(\lambda) d\lambda}. \quad (2.12)$$

2.3.2 Photometric units and standards

The term ‘light’ historically refers to that part of the electromagnetic spectrum visible to the human eye, in accordance with the data presented in Table 2.1. Light is sometimes referred to in a wider context with the term visible light emphasizing a confinement to wavelengths between 390 and 770 nm. Photometry is the

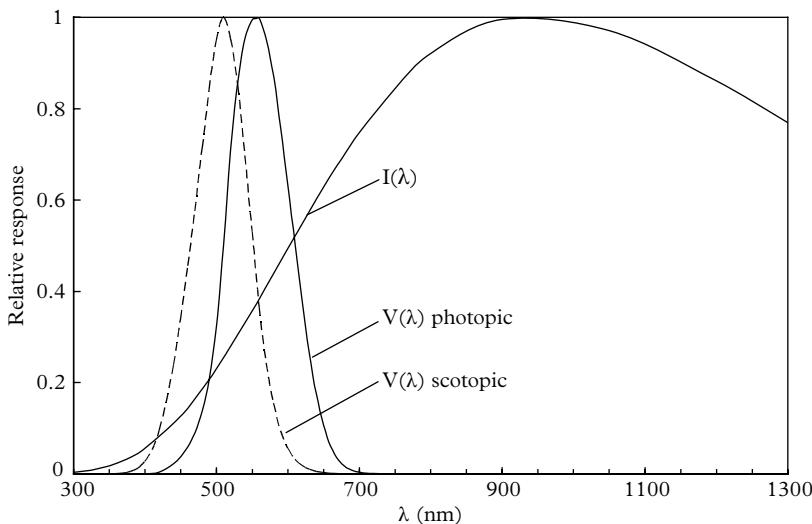


Fig. 2.2. Relative luminous efficiency functions $V(\lambda)$, together with the spectral radiance of a standard tungsten lamp, $I(\lambda)$. Note the limited overlap of the $I(\lambda)$ and $V(\lambda)$ curves.

science of measuring visible light in units that are weighted in accordance with the response of the human eye, illustrated in Fig. 2.2. Radiant energy falling outside the visible spectrum does not contribute to photometric quantities. The human eye was first officially calibrated as a detector in 1924, by averaging the visual response of over 100 observers, resulting in the relative luminous efficiency relationship, known as the CIE (Commission Internationale d'Eclairage) photometric curve, $V(\lambda)$. The human eye responds differently when daytime adapted (photopic vision) compared with darkness adapted (scotopic vision). Photometric measurements are based on the photopic curve that peaks at 555 nm, while the scotopic curve peaks at 500 nm.

Given the spectral radiant power of a lamp, shown normalized in Fig. 2.2 and designated $I(\lambda)$, one can calculate the luminous sensitivity S of a photocathode, if the QE distribution, $\eta(\lambda)$, is known; S is determined by reading the photocathode current generated by a tungsten lamp at 2856 K and dividing this quantity by the luminous flux incident on the photocathode. The lumen represents a relatively bright source and a flux of 1 millilumen is more appropriate for calibration purposes; the resulting photocurrent is

$$I_k = \left(\frac{C}{1.24} \right) \int_0^{\infty} I(\lambda) \lambda \eta(\lambda) d\lambda \quad \mu\text{A}, \quad (2.13)$$

where C is a constant dependent on the geometry and the intensity of the source. The luminous flux from the source, F, is given by

$$F = 683 C \int_0^{\infty} I(\lambda) V(\lambda) d\lambda \text{ lm},$$

$$S = \frac{I_k}{F} = \frac{10^3 \int_0^{\infty} I(\lambda) \lambda \eta(\lambda) d\lambda}{1.24 \times 683 \int_0^{\infty} I(\lambda) V(\lambda) d\lambda} \text{ } \mu\text{A/lm}, \quad (2.14)$$

using the definition of the candela as the luminous intensity of a source of frequency 5.4×10^{10} Hz, or equivalently 555 nm, with radiant intensity 1/683 W/steradian. Note in (2.14) that relative $I(\lambda)$ values may be conveniently used in the calculation, since this function appears in both lines of the equation. This calculation cannot be done in reverse: that is, it is not possible to deduce $\eta(\lambda)$ given S ; even the average value of $\eta(\lambda)$ cannot be deduced from a knowledge of S . Furthermore, S is a poor guide to the maximum QE, because the form of $I(\lambda)$ makes the luminous efficiency sensitive to the presence of any long wavelength tail in $\eta(\lambda)$. A quality bialkali photocathode may have a relatively low luminous sensitivity of 50 $\mu\text{A/lm}$ and yet a peak QE of 30% at 400 nm. An S20 photocathode with a QE of 15% at 400 nm may have a sensitivity as high as 250 $\mu\text{A/lm}$; the sensitivity is boosted by contributions to $I(\lambda)\lambda$ in the numerator of (2.14) generated by the long wavelength content of the spectrum.

The availability of standard light sources is a prerequisite for quantifying the sensitivity of any light detector. The luminous intensity (traditionally referred to as candle power) is the luminous flux per unit solid angle in a specified direction in units of lumen per steradian (lm/sr), or candela. The candela is defined historically in terms of the radiation from a black body at the temperature of platinum solidification; a candela is 1/60th of the luminous intensity of 1 cm² of such a source. As with radiance standards, international laboratories perform inter-comparison of photometric units (Ohno 1999). These arcane light sources are unlikely to be found in any establishment other than a standards laboratory. The development of a more convenient standard illuminant followed in the form of a tungsten filament lamp of a specified design and operated at 2856 K; clearly better suited to more general use. Standards laboratories can provide calibration as described by Ohno (1997). Ultra-bright white LEDs have an increasing role in industrial and home lighting and NIST (National Institute of Standards and Technology; previously known as the National Bureau of Standards of the USA) offers a calibration service for these products.

2.3.3 Filter measurements

It is so far clear that S is not necessarily a good selection parameter for applications other than white light. However, its usefulness can be extended for specific applications by interposing colour filters between a PMT and a standard white light source. The majority of PMT applications fall into three light categories: blue, red, and infrared. Photocathodes are individually calibrated for particular applications by noting the response to filtered light from a white light source. If the spectral transmission function of the filter is $f(\lambda)$, the response is given by including this function in the numerator of (2.14); thus,

$$S_f = \frac{I_f}{F} = \frac{\int_0^{\infty} f(\lambda) I(\lambda) \eta(\lambda) d\lambda}{1.24 \times 683 \int_0^{\infty} I(\lambda) V(\lambda) d\lambda}. \quad (2.15)$$

Given a QE curve, $\eta(\lambda)$ versus λ , and the filter function, $f(\lambda)$, (2.15) can be numerically integrated, using the data in Fig. 2.2 and Fig. 2.3, resulting in the entries in Table 2.2(a) and (b).

A Corning CS-5-58 filter, of half stock thickness is the basis for the so-called Corning Blue specification, designated by a CB number. The benefit of a reduced thickness is an increase in the light transmitted, while marginally altering the transmission spectrum. The CB sensitivity CB, quoted on a test ticket, is the photocathode current measured with the filter in place while maintaining the same operating conditions that apply when measuring S . The range in CB is

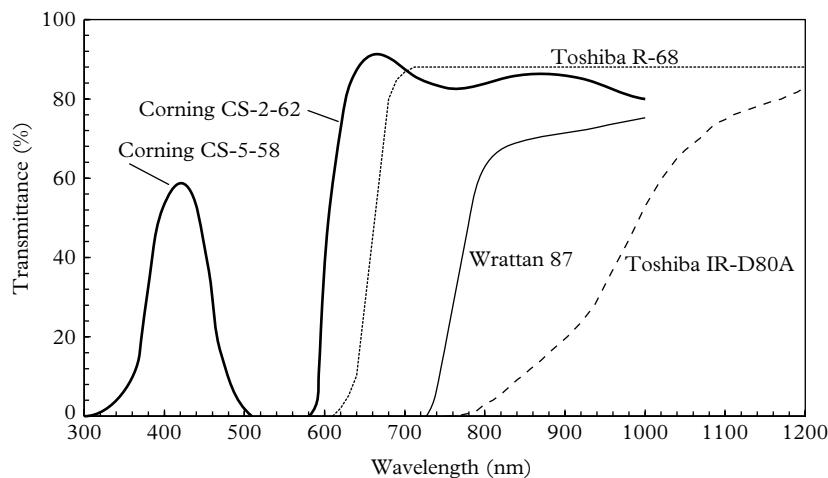


Fig. 2.3. Filters used by manufacturers.

34 Photocathodes

Table 2.2(a) Typical photocathode responses for end window PMTs offered by ET Enterprises.

Photocathode	S ($\mu\text{A/lm}$)	CB	CR	IR
K ₂ CsSb	50	12	1.5	—
Rb ₂ CsSb	110	12	10	—
Na ₂ KCsSb	250	10	140	20

Table 2.2(b) Filter- and laser-based measurements adopted by Hamamatsu Photonics. The red and infrared responses are chosen at well-known laser wavelengths. The R3896 is a side window PMT, and the remaining entries have a slip of activated crystal, as in Fig. 2.23(k).

Photocathode	S ($\mu\text{A/lm}$)	CB	R/W	633 nm	852 nm	1000 nm	mA/W
S20, R3896, SW	525	15	0.4	73	5	0	
GaAsP, R943	600	—	0.6	70	70	0	
InGaAs(Cs), R3310	150	—	0.4	22	10	2	

8 to 15, and a rough empirical conversion to the peak QE % at 420 nm is provided by the product of $CB \times 2.5$. CB is a useful selection parameter for the majority of scintillators, in particular NaI(Tl) and plastic. More recent materials such as LaBr₃(Ce) peak at 380 nm but the CB number is still relevant for grading or specification purposes. Sometimes the CB-filtered response is clumsily referred to as microamps per blue lumen, but it is best to avoid this description.

A Corning CS-2-62 filter defines the Corning Red response CR , which is indicative of sensitivity beyond 600 nm. The red:white ratio is a similar measure to CR but is based on the response to a Toshiba R-68 filter that cuts in at 680 nm. The infrared parameter IR is determined with a Wrattan 87, or Toshiba IR-D80A filter, intended for the selection of S20 and III-V photocathodes.

The response to a Corning filter CS-5-58, of half stock thickness, is accepted throughout the industry but this does not generally apply to the other filters. Response figures may be measured using a white light source, or, if the spectral response is known, they can be deduced by inserting $f(\lambda)$ in (2.15). There is a relationship between the red and infrared sensitivity of a photocathode and its dark current; photocathodes characterized by high red and infrared filter numbers tend to have high dark current. Filter numbers are thus indicators of the expected PMT dark current.

2.3.4 Calibration laboratories and absolute radiation standards

Radiometry is the science of measuring electromagnetic radiant energy. Spectro-radiometry is concerned with the spectrum of the radiating source, usually confined to a narrow spectral band. The goal of radiometric calibration is to quantify the relationship between the current output of a photodetector and the incident radiation flux. High accuracy cryogenic radiometers (HACRs) are the absolute standards primarily for the measurement of radiation flux; they are based on laser wavelengths spanning 406 to 920 nm, and subject to 0.02% (3σ) uncertainty. HACRs have been adopted by several national laboratories as primary standards for generating more convenient transfer standards, as described in the informative reports by Larason *et al.* (1996) and Zheng *et al.* (2000). It is encouraging to note that international comparisons of photodetector responsivity measurement have been commissioned from time to time and the results published; for example, Köhler *et al.* (1995) report favourably on the unanimity of a collaboration of 18 standards laboratories in the measurement of spectral responsivity. Overall agreement applies over the range 250 to 1000 nm for silicon detectors, but with noticeable disagreement in the UV region of the spectrum.

The operating principle of an HACR is electrical substitution radiometry (ESR), in which the heating effect produced by the absorption of incident radiation is matched to that required by electrical heating in order to produce the same power. This highly developed instrument is capable of measuring to an accuracy of 0.02% over a range of wavelengths. Measurement may take days to execute, demanding quicker ways for calibration. Practical means, based on secondary standards, have been devised to transfer the optical power scale of HACRs to more convenient detectors. A monochromator-based spectral comparator facility (SCF) provides a continuous wavelength scale of absolute spectral power response; the measuring instrument is trap detector, which is described in 2.3.5. By this means, the wavelength coverage can be extended to 200 nm but the uncertainty ranges from 0.22% to as high as 3.5%, as shorter wavelengths are approached. The information presented here is based on the systems described by Larason *et al.* (1996) working at NIST. It is recommended that users who need calibrated detectors should first consult their national standards laboratory, as major industrial countries offer similar calibration facilities to NIST.

2.3.5 Trap detectors

Silicon photodiodes have attributes such as long life, wide dynamic range, ruggedness, and high QE, making them ideal as detector standards. However, the realizable QE of an individual diode is about 85% because of reflection losses. These losses stem from the high refractive index, of about 3 for pure silicon. In practice, a 30 μm film of low-refractive-index SiO_2 is applied to the surface of a

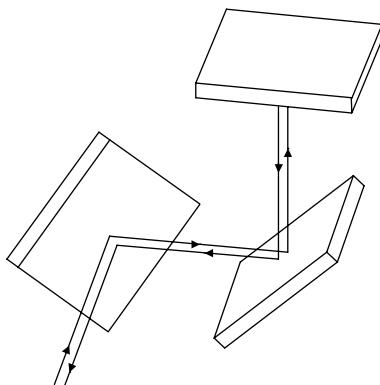


Fig. 2.4. The five-pass trap detector based on three silicon diodes connected in parallel.

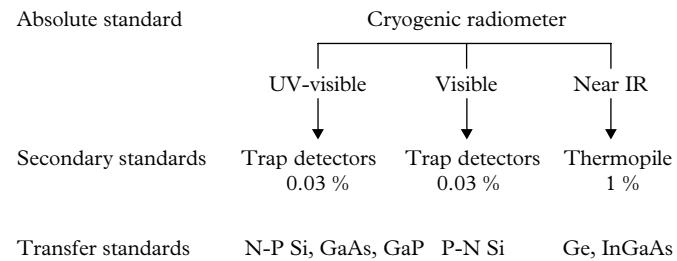


Fig. 2.5. Transfer chain of an absolute standard to a detector available to users. A representative set of detectors for transfer standards is listed in the bottom line.

photodiode to reduce the losses, which still requires considerable correction. A silicon trap detector for collimated light, devised by Zalewski and Geist (1980), absorbs essentially all incident radiation. This is achieved in a multiple pass arrangement of a set of three windowless silicon photodiodes assembled in a non-planar geometrical arrangement, shown in Fig. 2.4.

Diodes are connected in parallel and the combined output current feeds a high-gain transimpedance amplifier. The optical path within the detector involves five reflections insuring an overall absorptance close to unity; a small, but well-defined correction, may also be applied to the deficit. Note the similarity in the light capture process to the multi-bounce optical enhancer discussed in 3.13.3. The original trap detector was based on a UDT UV100 diode but this has been superseded by a Hamamatsu PN S1337 diode. An example of the application of a trap detector for generating transfer standards from a cryogenic radiometer is shown in Fig. 2.5.

The range from 200 to 1000 nm is of prime importance in PMT applications, and standards laboratories, such as NPL in the UK, can provide a spectral responsivity calibration, in 10 nm steps, for a PMT photocathode. The PMT is

operated as a vacuum photodiode with light incident normally and covering a specified area of the photocathode. This calibration is compromised whenever a PMT is operated under gain, as intended. In particular, if light is incident over a range of incident angles covering the entire sensitive area of the photocathode. These sources of uncertainty, discussed in 2.4.5, seriously affect the ability of PMTs in making reliable absolute measurements of light flux. There is a further source of uncertainty related to the controversial subject of collection efficiency, explored in 10.2.

2.4 Optical properties of end window photocathodes

2.4.1 Refraction in dielectrics

Any system, consisting of elements with the same refractive index, may be optically coupled without loss in transmission at the interfaces. Optically coupled implies that all components are joined by a transparent film of similar refractive index (usually a viscous silicone compound). Transmission losses occur through reflections when elements have dissimilar refractive indices. Most inorganic scintillators, characterized by high Z , have a refractive index exceeding that of glass (e.g. 1.85 for NaI(Tl)). This leads to light trapping in the corners of scintillators of cylindrical geometry, with little prospect of release (Fraser-Mitchell and Wright 1990). On the other hand, liquid and plastic scintillators closely match the refractive index, n , of PMT window glasses. All such materials are characterized by $n \sim 1.5$ with little light loss when optically coupled. The reflection coefficient for light in transit across a dense to less dense interface is theoretically 100% for incident angles beyond the critical one (Fig. 2.6(b)). Total internal reflection (TIR) is turned to advantage in the practical realization of optical enhancement and light piping, as covered in Chapter 3.

The refractive indices of a PMT window and the photocathode have a fundamental role in the efficient collection of light. Photocathodes intended for end window devices are semi-transparent, thin, and described by a complex refractive index $\mathbf{n} = n + ik$. The optical characterization of photocathodes is covered in 2.4.3 and 2.4.6, but for immediate discussion, it is sufficient to note that $n \sim 2.7$ and $k \sim 1.5$ for bialkali photocathodes. The imaginary component, k , is responsible for absorption, and it is related to QE; the real part of the refractive index, n , is a familiar parameter in classical optics. Initially, we consider only this component for the study of reflections, discovering that a high numerical value for n leads to unexpected consequences in PMT behaviour.

According to Snell's law, light passing at an angle θ_1 , from a medium of refractive index n_1 to another with a coefficient of n_2 , is refracted by an angle θ_2 given by

$$\theta_2 = \sin^{-1} \left[\left(\frac{n_1}{n_2} \right) \sin \theta_1 \right]. \quad (2.16)$$

Following Gunter *et al.* (1970), for layered media with refractive indices n_1, n_2, \dots, n_j ,

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 = \dots = n_j \sin \theta_j = C, \quad (2.17)$$

where C is a constant. If the first and last media are the same, $n_1 = n_j$, then, from (2.17), $\sin \theta_1 = \sin \theta_j$, prescribing that the final ray must be parallel to the incident one. This is illustrated in Fig. 2.6(a). Light undergoes TIR whenever a medium of index $n < C$ is encountered. The onset of TIR is signalled by $C = 1$, and hence $\theta_1 = \sin^{-1}(1/n_1)$ (see Fig. 2.6(b)). There are certain consequences embodied in (2.17) which may not appear obvious. The angle at which a ray of light, with a given incident angle, emerges after crossing a series of parallel layers, of different refractive indices, depends only on the refractive indices of the first and last media. If the two indices are the same, the light emerges at the same angle as the incident one but laterally displaced, as shown in Fig. 2.6(a).

For our purposes, air and vacuum have a refractive index of unity. When PMTs are used in air, the incident light follows trajectories of the form shown in Fig. 2.6(a). Figure 2.6(b) illustrates the phenomenon of TIR, whereby transmission is replaced by reflection at the final interface. TIR cannot occur, regardless of the incident angle, in configurations such as Fig. 2.6(a).

There are applications, particularly in high-energy physics and astrophysics, in which light is generated in media other than air: such as water, ice, liquid argon, and a range of scintillators. TIR is possible in all these cases, and its onset is

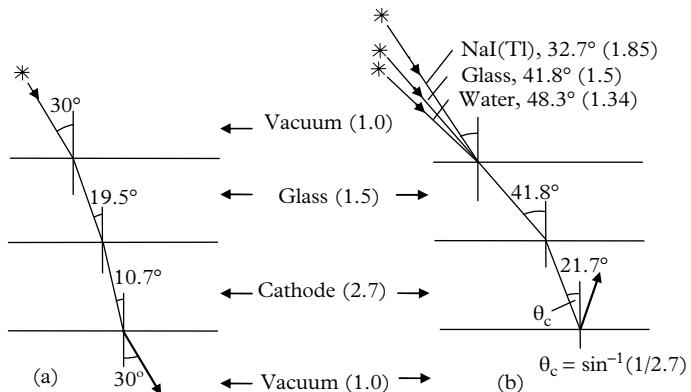


Fig. 2.6. (a) Ray paths showing that entrance and exit angles for layered media are identical, but with a lateral shift, when the first and last media have the same refractive index (shown in parenthesis). (b) Illustrates TIR at the cathode–vacuum interface, which can only occur if the first and last media have different refractive indices. In all cases, as indicated, incident light is created within the first medium.

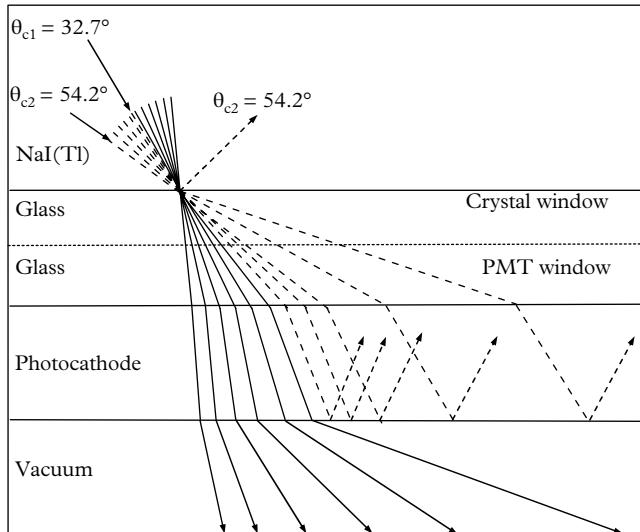


Fig. 2.7. Propagation of light generated within a NaI(Tl) scintillator. The dashed trajectories refer to paths that involve TIR at the cathode–vacuum interface; θ_{c1} and θ_{c2} are critical angles for TIR. Reflections (not shown) occur at all interfaces and for all incident angles in accordance with the Fresnel formulations discussed in 2.4.2.

illustrated for water, glass, and NaI(Tl) in Fig. 2.6(b). Note that TIR occurs at the cathode–vacuum interface, and the incident angle at which this occurs is determined by the refractive indices of these two media only. The escape of light from a high-refractive-index material, such as sodium iodide, is examined in closer detail to illustrate the difficulties in extracting light from an encapsulated cylindrical crystal. That is, one coated in diffuse reflector, with the exception of one plane face through which scintillator light is permitted to exit. Referring to Fig. 2.7, note that all light leaving the crystal at an incident angle of 32.7° enters the vacuum unless absorbed by the cathode; for angles between 32.7° and 54.2° , light is either absorbed or reflected off the cathode–vacuum interface and returned to the crystal. At incident angles $>54.2^\circ$, light is reflected at the NaI(Tl)–glass interface and returned to the crystal. Reflected light is not necessarily lost in a quality crystal, that is, one of long attenuation length and clad with an efficient reflector.

If the window of an encapsulated scintillator is optically coupled to that of the PMT, TIR does not occur at the window–cathode interface but beyond a certain angle it occurs at the cathode–vacuum interface. The proportion of light transmitted through the photocathode (ignoring absorption) plus that totally internally reflected from the cathode–vacuum interface is given by the solid angle formula (2.1): $\Omega = \frac{1}{2}(1 - \cos 54.2^\circ) = 21\%$. If an air gap is introduced between the two windows, light incident at angles greater than $\sin^{-1}(1.0/1.85) = 32.7^\circ$ is totally reflected at this glass–air interface, and the solid angle for light collection is

reduced to $\Omega = \frac{1}{2}(1 - \cos 32.7^\circ) = 8\%$. The total light collected in an ideal optical arrangement, including multiple passes but ignoring attenuation and reflection losses, is the same for both solid angle considerations. However, the significant difference between these two single pass collection percentages points to the advantages of optical coupling between the scintillator and PMT window in any practical configuration in which light is lost by absorption, and leakage at the periphery of the crystal–PMT interface; this can be as much as 30% and is discussed in 3.4. Note that at a NaI(Tl)–glass interface the transmitted angle is always away from the normal, further exacerbating light leakage. A useful and easily executed exercise on light loss is to record a pulse height distribution with and without optical coupling between a crystal and a PMT window. Non-optical coupling may be achieved by standing-off the crystal by a fraction of a millimetre using three small diameter studs. Experimentally, it is found that only about half the light is collected from a NaI(Tl) crystal under these circumstances, compared with the adoption of optical coupling. Manufacturers of inorganic scintillators appreciate the benefits of a thin window for encapsulation, or no window as in an assembly. The advantage offered by air coupling is one of long-term stability, because coupling fluid has been known to migrate over time. These considerations are further explored in Chapter 3. Light collection from a plastic scintillator, which has approximately the same refractive index as glass, is clearly more efficient, although the intrinsic light yield is considerably lower than for the inorganic scintillators. This treatment highlights the difficulties in capturing light generated in an optically dense medium. Furthermore, light collection is actually smaller than presented because specular reflections occur at all interfaces; although these are small in amplitude, except at glancing incidence. The discussion so far is limited to tracing the path of propagated light without taking account of inter-media reflections. An exact formulation for this is given by Fresnel's treatment of Maxwell's equations, given in 2.4.2.

2.4.2 Reflection from dielectrics

Arguments centred on TIR in 2.4.1 reveal paths for returning light to a crystal. However, partial reflection of incident radiation occurs at every optical interface, making a contribution to returned light. Fresnel's theory for dielectric materials quantifies the process, and is covered in any good book on optics. The standard text is that of Born and Wolf (1999), but it is not for the faint-hearted. There are two plane-polarized components to consider: perpendicular, described by s or TE, and a parallel component referred to by p or TM; r_s and r_p are the respective reflectances given by

$$\begin{aligned} r_s &= \left[\frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)} \right]^2, \\ r_p &= \left[\frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \right]^2, \end{aligned} \quad (2.18)$$

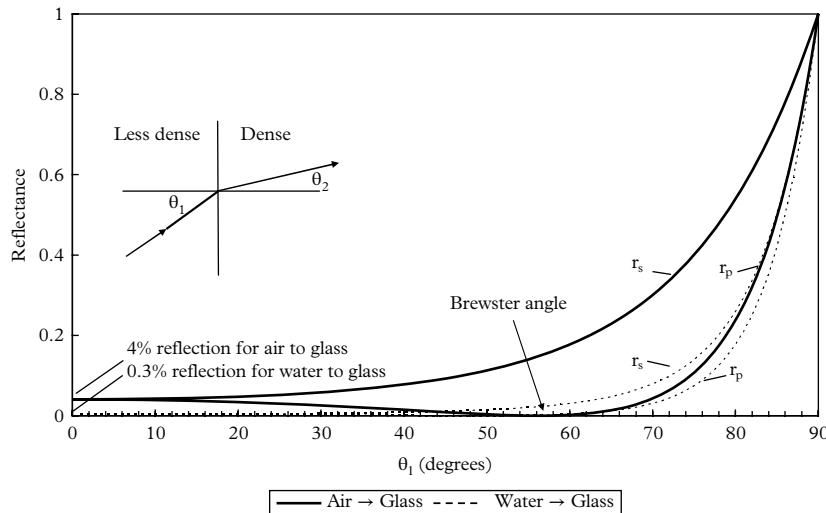


Fig. 2.8. Reflectances at the transition from a less dense to a more dense medium, according to Fresnel's theory.

where θ_1 and θ_2 are the incident and refracted angles, respectively; they are related by Snell's law, stated in (2.17). The formulations of (2.18) are indeterminate at normal incidence but the equations in this case can be extracted by using the identities $\sin(\theta_1 - \theta_2) = \sin\theta_1\cos\theta_2 - \cos\theta_1\sin\theta_2$, and setting tangents equal to sines, to arrive at

$$r_0 = r_s = r_p = \left(\frac{n - 1}{n + 1} \right)^2, \quad (2.19)$$

where n represents the ratio of refractive indices at the interface of any two media. Equation (2.19) predicts that 4% of normally incident light is reflected at a glass-air interface (strictly, glass-vacuum). Replacing n by $1/n$, gives the same outcome for r_0 , confirming that the reflected intensity is invariant to direction for light normal to an interface: the total light reflected is therefore 8% for two interfaces. Interchanging θ_1 and θ_2 confirms the general directional invariance of r_s and r_p . The angular dependence is shown in Fig. 2.8 for air-glass and water-glass interfaces. The reflected component is small for water-glass transitions, for all incident angles of up to $\sim 60^\circ$, indicating that a perfect match is not essential when coupling two different media, provided they are reasonably close in refractive index.

The shape of the curves is a common feature in all transparent materials, showing a steep increase as θ_1 approaches 90° . The formulae predict that all smooth dielectric surfaces become highly reflective at glancing angles: an everyday experience with reflections in glass, water, and even non-dielectric materials. This also applies to many diffuse reflectors, such as PTFE and TiO_2 , that in fact exhibit

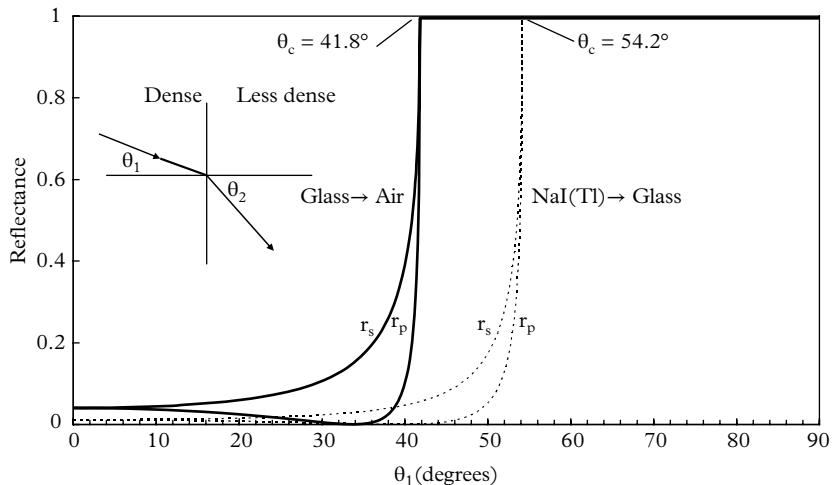


Fig. 2.9. Reflectances at the transition from a dense to a less dense medium. Note the rapid rise in r as TIR is approached. The critical angle for NaI(Tl)-glass is 54° , and the reflectance for light incident normally is 1.1%.

specular reflection at incident angles beyond 70° . According to Janecek and Moses (2008a, b) TiO₂ paint provides the closest to a Lambertian distribution of all the diffuse reflectors so far studied. This has serious implications for Monte Carlo simulations unless a combination of specular and Lambertian is assumed. In the reverse case, of dense to less dense transitions, the shape of the curves is similar until TIR occurs, as shown in Fig. 2.9.

The reflection coefficient, at any incident angle for s polarization, is everywhere higher than for the p mode. The behaviour of reflection coefficients for p polarization is complicated because of the influence of the two Brewster angles, at 34° and 56° , determined by $\tan^{-1}\theta_c$ for glass-air and air-glass, respectively. High reflectance, $r = 0.21$, is predicted for a photocathode with $n = 2.7$ from (2.19). Photocathodes are known to have a thickness in the region of 20nm, with measured values of r in agreement with the figure of 21%. According to Moorhead and Tanner (1996), the reflection process in a thin film can be understood in terms of multiple reflections within the photocathode that interfere coherently with one another. These amplitudes sum to determine the total amplitudes of the reflected and transmitted waves.

2.4.3 Photocathode optical properties

The only known research into photocathode parameters published by an industrial organization is that due to Timan (1976) while working for DuMont Electron Tubes and Devices Corp (the company ceased trading in the late 1970s).

Ambiguities in Timan's treatment of experimental results for n , as a function of wavelength, have been resolved by Moorhead and Tanner (1996) who re-presented Timan's results for a bialkali photocathode. These latter authors, working with Thorn-EMI Electron Tubes, made angular measurements at a fixed wavelength of 442 nm, while those of Timan are restricted to normal incidence but cover a range of wavelengths. The motivation for these studies is quite different, and interesting. Timan's aim was to use knowledge of the optical characteristics, measured during activation, as a means for making higher efficiency photocathodes. Moorhead and Tanner (1996), and Lay (1996), through their participation in the development of the Sudbury Neutrino Observatory (SNO), were concerned with mimicking the operation of some 10,000 PMTs immersed in a massive tank of water. For this reason, their investigations were performed in the same medium, and considerations such as polarization, variation of QE with incident angle, and reflectance of bialkali photocathodes were essential to achieving a reliable Monte Carlo simulation of their detector. A group led by P D Townsend was primarily concerned with enhancing the QE of S20 PMTs. Covering a wide range of wavelengths, they provided data (see Fig. 2.15) on optical constants for S20s, where previously such detail was meagre (Hallensleben *et al.* 1999, 2000).

2.4.4 Techniques for measuring photocathode reflectance

One of the earliest determinations of photocathode parameters is that of Jones (1976) using a half-cylindrical lens in optical contact with a PMT window, shown in Fig. 2.10(a). Four per cent of incident light is lost at the air–Perspex interface but, importantly, this loss is the same for all angles of incidence and cancels in ratio-metric computations. Jones was able to deduce photocathode reflectivity alone, by the ratio of photocurrents, with and without the aluminium reflector shown in the figure. Measurements were made at 420 nm, 560 nm, and 694 nm for both polarization states based on an EMI 9558B, S20 PMT. The angular sensitivity response—the measured anode current—does not follow Jones' theoretical predictions; possibly because anode current is proportional to QE, which is not synonymous with absorptance—the latter always exceeds QE in magnitude. Jones' measurements show a decreases in sensitivity with incident angle up to the critical angle for glass, after which enhancement is evident. The sensitivity doubles for s polarization, at 694 nm, compared with normal incidence, in quantitative agreement with the computations of Hallensleben *et al.* (2000) who determined a factor of 8 increase at 750 nm for S20 photocathodes. The measurement technique employed by Hallensleben *et al.* is similar to that of Jones but with the inclusion of an independent viewing PMT, in preference to measuring photocurrent in the PMT under consideration. This is the mode of operation adopted by Moorhead and Tanner (1996) and Lay (1997) working in the same laboratory.

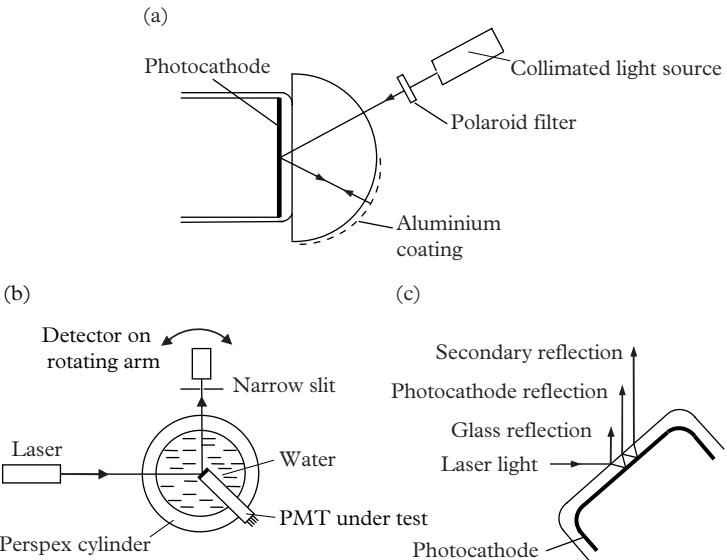


Fig. 2.10. Conceptual illustrations of apparatus used for measuring photocathode parameters.

(a) Based on a half-cylindrical Perspex lens

Jones, D. P. (1976). Photomultiplier sensitivity variation with angle of incidence on the photocathode. *Appl. Opt.*, 15, No. 4, 910–14.

(b) Immersed in water within a full cylindrical lens

Moorhead and Tanner (1996). Optical properties of an EMI K₂CsSb bialkali photocathode. *Nucl. Instr. and Meth. in Phys. Res. A*, 378, 162–70. Reprinted from Elsevier Science ©1996.

(c) Multiple reflections off two window faces.

The interesting feature of this apparatus in Fig. 2.10(b) is the choice of water as the medium for incident light, thus mimicking actual environmental conditions of the SNO detector, with an added advantage of reduced interface reflection losses. The laser beam shown produces multiple reflections within the glass window, which are separated spatially by the action of a 2 mm-thick PMT window. The width of the laser beam is 1.1 mm, thus allowing isolation of the different orders of reflection, as is evident in Fig. 2.10(c).

2.4.5 Limits to reliable measurement

It is desirable to measure all the optical parameters of a photocathode but it transpires that reflectance is the most amenable. Photocathode sensitivity can be determined either with the PMT fully operational or configured as a photodiode but in both cases the measured quantity is anode current. There are uncertainties in relating sensitivity to QE, or to absorptance, for the following reasons: the

cylindrical section of envelope between the window and the input to the dynode stack is aluminized and photosensitive, with a response similar to the photocathode proper (that is the area of cathode covered by the window). Electron-optics of PMT front-end designs are such that a fraction of the photoelectrons released from a sidewall are collected and thus contribute to anode current, whether operation is as a photodiode or PMT. A further anomaly is reflection off parts of the multiplier structure: most PMTs include a polished locating flange that faces the photocathode and returns a portion of incident light (see Fig. 2.33). This contribution is readily observed by flying-spot examination whereby a dip of $\sim 10\%$ in the response is seen, corresponding to the size and location of the aperture. The first dynode is photosensitive in most PMTs and it contributes to anode current but its contribution is reduced by a factor of $1/\delta_1$, where δ_1 is the gain of the first dynode. There is no contribution from this source when a PMT is operated as a diode, since all dynodes are connected to the anode in this instance.

It is clear from this that measuring absorptance directly is challenging. However, the difficulties disappear whenever measurements are made under TIR, for then no light enters the PMT beyond the photocathode. Under these circumstances, it is possible to relate QE to measured absorptivity. Lay (1997) has carried out such an exercise and finds close agreement under TIR between theory and experiment. Corrections for internal reflections are necessary at angles below the critical and agreement is restricted to verifying that photocurrent and absorptivity are proportional for all incident angles—this is an important finding.

2.4.6 Determination of transmission and absorptance

A photocathode may be characterized by three parameters: its thickness and the real and imaginary parts of the refractive index, n . The response to the two planes of polarization, p (parallel, TM) and s (perpendicular, TE) then follows. Reflection and transmission at the interface of dielectrics, such as glass and water, are described by the laws of Fresnel, as shown in Figs 2.8 and 2.9. The inclusion of a photocathode adds complications in the determination of optical constants: it is a thin, semi-transparent, semiconductor film with a thickness, d , of only ~ 20 nm. Consequently, because of absorption, the refractive index, $n = n + ik$ is complex and the Fresnel equations must be modified in the manner described, for example, by Lay (1996). The method adopted by several authors is to make optical measurements on photocathodes at multiple angles, and multiple wavelengths for p and s polarized light. The data allow determination of the three constants, n , k , and d , which cannot be calculated directly from experimental data. A three-parameter best fit is sought that leads to the most likely set of cathode parameters. As a refinement, Moorhead and Tanner (1996) used a four-parameter fit by including the refractive index of the PMT window into their fitting routine. Using reflection data, an example of which is shown in Fig. 2.11,

46 Photocathodes

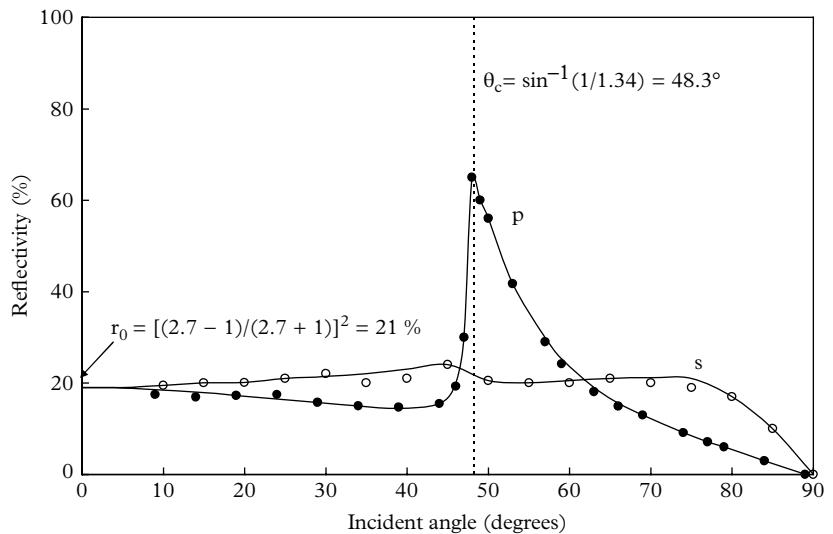


Fig. 2.11. The results of the best-fit procedure, solid line, and actual measurements at 442 nm yielding $n = 2.7 + 1.5i$, and $d = 23$ nm for a 9124B PMT operated under water. Fitting was performed on the p data.

Moorhead and Tanner (1996). Optical properties of an EMI K₂CsSb bialkali photocathode. *Nucl. Instr. and Meth. in Phys. Res. A*, 378, 162–70. Reprinted from Elsevier Science ©1996.

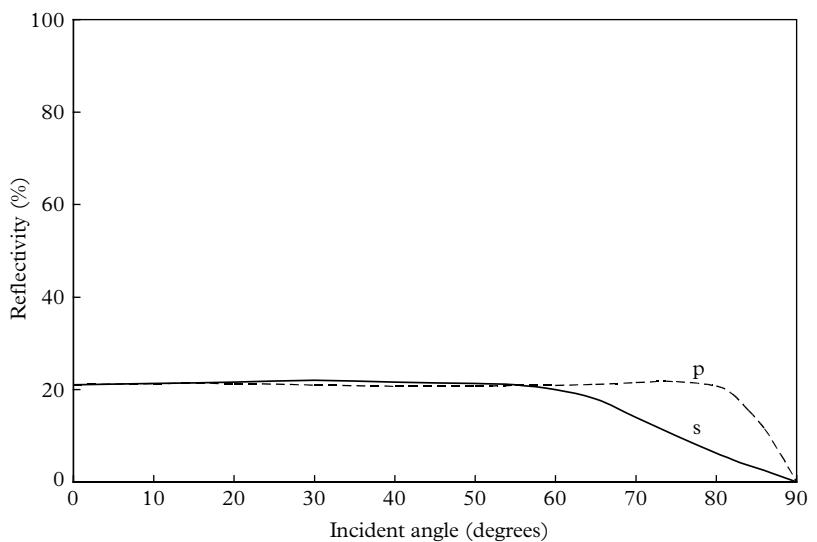


Fig. 2.12. Reflectivity as a function of incident angle with the PMT window in air.

Moorhead and Tanner (1996). Optical properties of an EMI K₂CsSb bialkali photocathode. *Nucl. Instr. and Meth. in Phys. Res. A*, 378, 162–70. Reprinted from Elsevier Science ©1996.

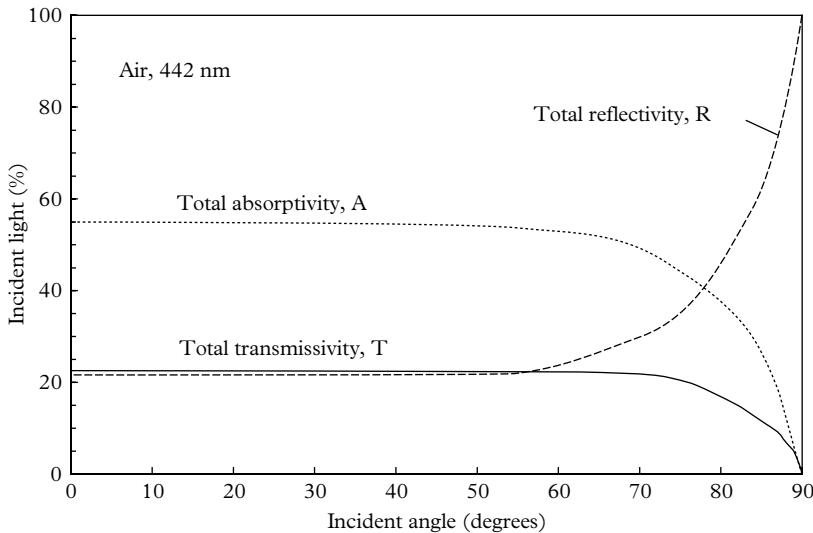


Fig. 2.13. The characterization of a 9124B PMT operated in air, in terms of three parameters: A, R, and T, which sum to 100% at all angles. Equal contributions are assumed for s and p polarizations.

Moorhead and Tanner (1996). Optical properties of an EMI K₂CsSb bialkali photocathode. *Nucl. Instr. and Meth. in Phys. Res. A*, 378, 162–70. Reprinted from Elsevier Science ©1996.

leads to $n = 2.7$, $k = 1.5$, and $d = 23$ nm for an EMI 9124B Rb₂SbCs photocathode immersed in water.

The relationship for reflectivity in air also follows by using the same optical constants, illustrated graphically in Fig. 2.12. Reflectivity results are provided by these authors for two samples of K₂SbCs photocathodes and one for a Rb₂SbCs photocathode. The three PMTs have similar characterization with respect to n and k , but with wider variation in d . The deduced values of the three optical constants lead to the determination of three entities: absorbance (A), reflectance (R), and transmissivity (T), subject to the condition $R + T + A = 1$, or $R + A = 1$ when TIR is present. Results due to Moorhead and Tanner are reproduced in Fig. 2.13, assuming equal contributions from s and p polarizations (that is unpolarized light) with the PMT window in air. Graphs are featureless, lacking the influence of TIR because the initial and final media are common (air and vacuum). A set of results is presented in Fig. 2.14, based on water as the initial medium. The highly structured responses for A and R bear the mark of TIR, which is also clearly evident in Fig. 2.11.

The studies by Moorhead, Tanner, and Lay (MTL) concerned a single wavelength of 442 nm and bialkali photocathodes. Motta and Schönert (2005) measured the absolute reflectance at near-normal incidence, together with the change in polarization upon reflection from the photocathode. Results were taken with an ellipsometer spanning 380 to 680 nm. Their fitting procedure assumed the MTL constants at 442 nm, that is, $n = 2.7$, $k = 1.5$, for K₂CsSb,

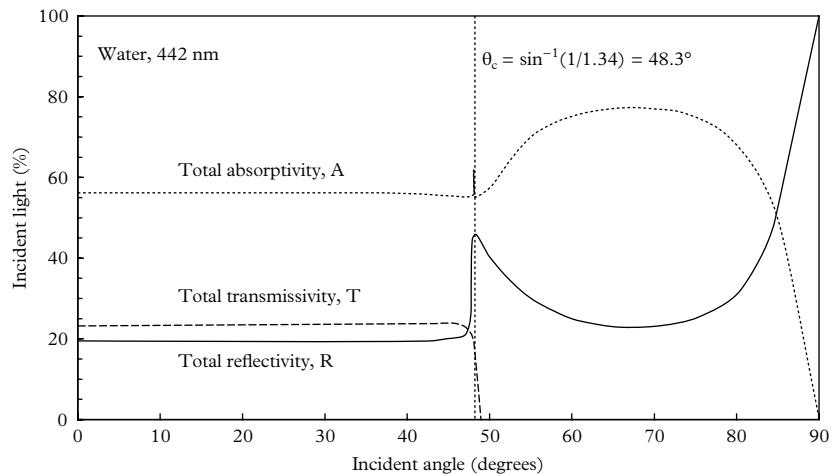


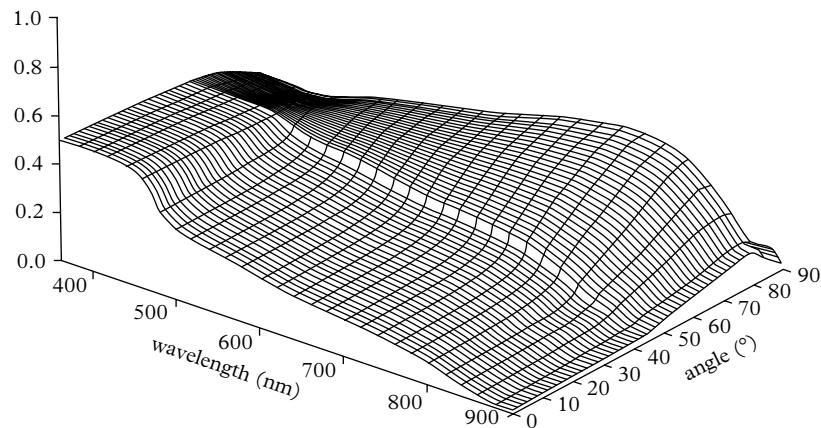
Fig. 2.14. The companion graph to Fig. 2.13 for a PMT immersed in water.

Moorhead and Tanner (1996). Optical properties of an EMI K₂CsSb bialkali photocathode. *Nucl. Instr. and Meth. in Phys. Res. A*, 378, 162–70. Reprinted from Elsevier Science ©1996.

and $n = 2.5$, $k = 1.35$, for the Rb₂SbCs photocathode. On this basis, they were able to formulate the set of n , k , and d parameters as a function of wavelength. Knowledge of these constants leads to the R , T , and A relationships of the type shown in Figs 2.13 and 2.14.

A group led by P D Townsend made a theoretical study of the refractive index of S20 photocathodes, covering 360 to 900 nm (Hallensleben *et al.* 2000); the refractive index attains a maximum of 3.63 at 420 nm. At wavelengths in the red and infrared regions, where this type of cathode is best suited, k becomes vanishing small for wavelengths approaching 900 nm. The implication of this is that infrared light will preferentially pass through the photocathode rather than produce photoelectrons, a phenomenon well known to users. A means for recycling this light through the photocathode several times will lead to enhanced detection that increases with the number of cycles. This process is clearly more effective for IR than for blue light, and means for achieving this are discussed in Chapter 3. The A , T , and R relationships, as a function of wavelength and incident angle, contain a wealth of information—easily understood through the clear isometric projections of Hallensleben *et al.* Absorptance distributions are shown in Fig. 2.15 for two polarization modes. The upper distribution, s , is the more significant one for an S20 photocathode operating in the red and infrared part of the spectrum; this is because of the clearly discernible hilltop in the response centred on 750 nm and 75°. The absorptance takes the shape of a ridge increasing monotonically in height from 470 nm to a broad maximum near (750 nm, 75°); this response is eight times higher than that at normal incidence. Note for blue light, absorptance is flat up to the critical angle and then decreases with increasing θ . Enhancement is present in the p mode at angles beyond the groove, signifying the onset of TIR, but without

Absorptance TE



Absorptance TM

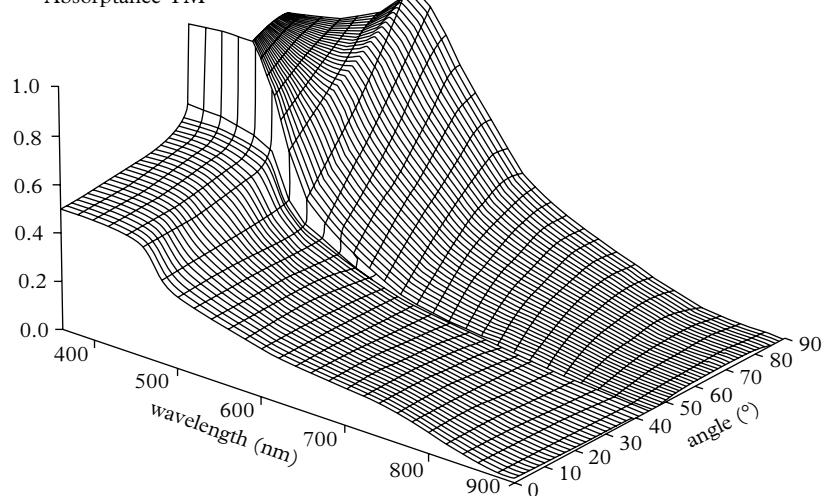


Fig. 2.15. Absorptance for TE(s) and TM(p) of an EMI 9657Q, S20 PMT. Absorptance for unpolarized light is the average of the two distributions.

Hallensleben *et al.* (2000). Optical constants for the S20 photocathode, and their application to increasing photomultiplier quantum efficiency. *Optics Communications*, 180, 89–102. Reprinted from Elsevier Science ©2000.

the maximum seen at 750 nm, 75°, in the s mode. The sharp knife-edge in the TM mode, situated on the critical angle and extending from 400 to 500 nm, is an intriguing feature that is evident in experimental results. There is significant enhancement in the p mode spanning 400 to 600 nm and peaking at 500 nm, at angles beyond the critical one.

Two considerations, based on the prediction that QE increases at glancing angles, and the adoption of light recycling, form the theoretical and practical basis of optical enhancement with S20 photocathodes. The same considerations apply to bialkali photocathodes and blue light except that absorptance is already high, as shown in Figs 2.13, 2.14, and 2.15, and there is thus limited enhancement with this photocathode type. Exploitation of the theoretical predictions is discussed in Chapter 3.

2.5 Spectral sensitivity of photocathodes

2.5.1 Three foremost photocathodes

Photocathodes have been known to survive for more than 30 years, depending on how the PMT is used, ambient light levels, and the state of the PMT vacuum. In addition to the relentless ingress of atmospheric helium, slow degassing of the internal metal and ceramic elements ultimately poisons the photocathode. Helium, being inert, does not react chemically with the photocathode, but the electron multiplication process generates helium and other ions, causing damage. An unusable gassy PMT is said to be 'soft'. Photocathodes are damaged by long-term exposure to bright lights. If an inactive PMT is continuously exposed to sunlight and fluorescent lighting in an office or laboratory, noticeable bleaching will take place over a time span of several months and the yellow/brown appearance of the photocathode ultimately disappears. A potential gradient between the window and photocathode of a PMT can cause migration of sodium ions from the glass to the photocathode, thereby poisoning it. This situation may arise in a PMT housing designed for use with negative high voltage (HV), should it include grounded metallic locating rings. Extremely low temperatures, approaching liquid argon, can be tolerated by most PMT types, although electrical performance may suffer. Operation at temperatures up to about 100 °C is permissible with standard bialkali PMTs, but the lifetime may be reduced. High-temperature bialkalies are usable up to 200°C, but serious loss of performance and reduced lifetime have to be accepted. These considerations apply to the more common types, but the III–V photocathodes need greater care in their use and storage—their high cost alone should demand this.

The majority of PMT applications, whether in research or industry, are served by just three photocathode types: bialkali (K_2CsSb), rubidium bialkali (Rb_2CsSb), and S20 (Na_2KsCs) with spectral responses shown in Fig. 2.16. Table 2.2(a) summarizes cathode sensitivity in terms of white light response, S ($\mu\text{A/lm}$), and the

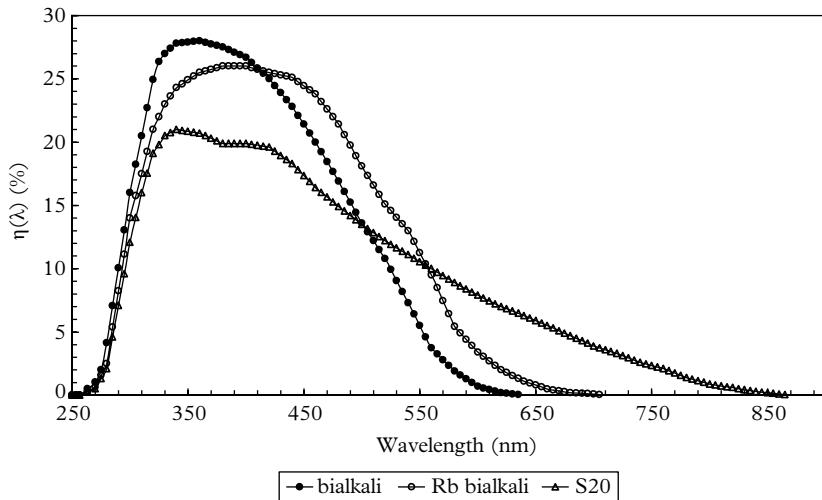


Fig. 2.16. Spectral sensitivity curves, $\eta(\lambda)$, for three widely used photocathodes. The rubidium photocathode clearly has higher QE for green light, compared with the potassium bialkali. The shoulder starting at 480 nm provides a boost to QE.

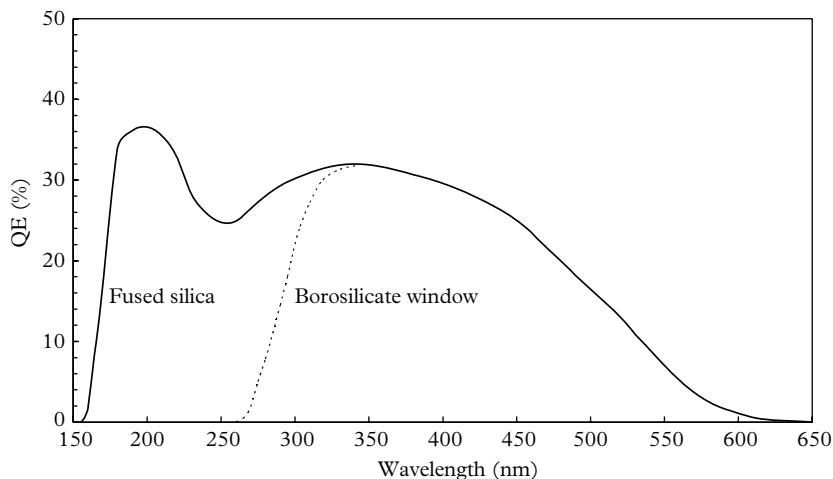


Fig. 2.17. Showing the extended response, for a bialkali photocathode, in the UV region of the spectrum with a fused silica window.

CB, CR, and IR numbers. The photocathodes depicted are all sensitive to UV light but this attribute is lost where the window is of borosilicate glass, for example. Detection of UV light requires a UV-transmitting window, such as of fused silica, to extend the spectral response, shown in Fig. 2.17. The other two common photocathode types behave similarly in the UV.

52 Photocathodes

These widely used photocathodes have a long history and until recently improvement in QE has been continuous but slow. Initially with a peak QE of ~15% at 400 nm for the three types, this figure had increased in the course of 50 years to about 25%, with the occasional photocathode exhibiting a peak QE of 30%. In 2005, Hamamatsu K K offered a highly sensitive metal package PMT with a so-called super bialkali photocathode (SBA) with 35% peak QE at 400 nm (Nakamura *et al.* 2010). This was followed in 2006 by the announcement of an ultra bialkali (UBA) with 45% QE at 400 nm. Furthermore, a shoulder near 500 nm ensures a QE of 30% at 500 nm, making this photocathode ideal for ATP (adenosine triphosphate) and blue-green scintillator applications. The UBA photocathode has the highest peak QE, among the range of available photocathodes, with the exception of the crystal type GaAsP photocathode, which has a QE of 55% at 550 nm.

2.5.2 Photocathode response to scintillator light

Three photocathodes fulfil the requirements for the detection of blue, green, and red light. There are different photocathode types for use in the vacuum UV (VUV) region of the spectrum and others suitable for infrared wavelengths. Given a light source of single wavelength, one simply selects the photocathode with the highest QE at that wavelength, but the decision is not straightforward where a spectrum of wavelengths is involved. Figures 2.18 and 2.19 illustrate the wavelength bands over which the more common scintillators emit light—the bands are typically 100 to 150 nm wide.

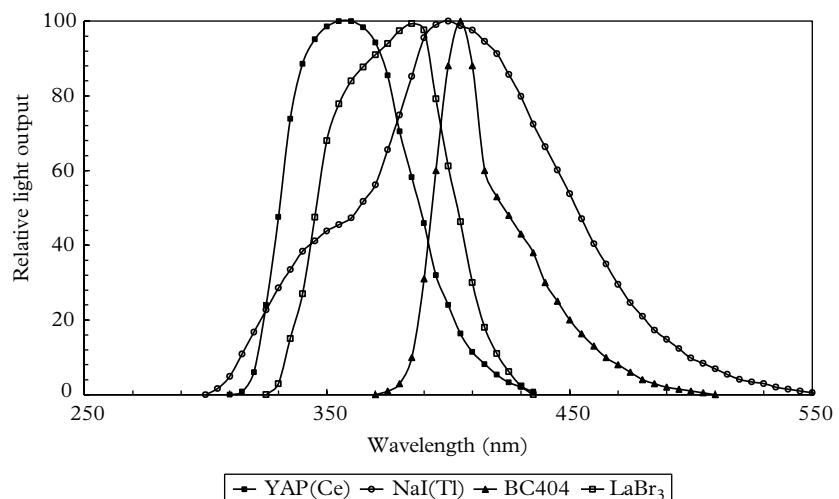


Fig. 2.18. Spectra for a representative range of scintillators emitting UV-blue light.

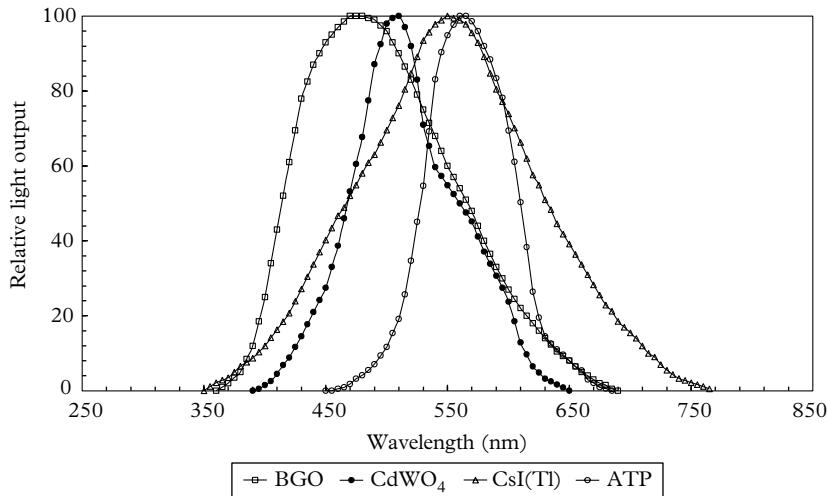


Fig. 2.19. Scintillator and ATP emission spectra for which green-red light is emitted.

Table 2.3 Mean QEs (%) allied to the emission spectra of selected scintillators, for three photocathode types. ATP (adenosine triphosphate) is included because of its importance in bioluminescence. Mean QEs are computed using (2.20).

Photocathode	LaBr ₃	YAP	NaI(Tl)	BC404	BGO	CdWO ₄	ATP	CsI(Tl)
K ₂ CsSb	27	27	25	25	14	11	4	8
Rb ₂ CsSb	26	25	25	25	18	16	8	11
Na ₂ KCsSb	20	20	19	19	14	13	10	11

The response of a photocathode to a spectrum of wavelengths is determined by computing the mean QE, $\bar{\eta}$, for the given spectrum, $\gamma(\lambda)$; thus,

$$\bar{\eta} = \frac{\int_0^{\infty} \eta(\lambda) \gamma(\lambda) d\lambda}{\int_0^{\infty} \gamma(\lambda) d\lambda}. \quad (2.20)$$

Equation (2.20) can be evaluated with sufficient precision by summation, with $\Delta\lambda$ taken at least every 10 nm. The entries in Table 2.3 are derived by applying (2.20) to a representative selection of scintillators, although the method can be applied to any light source.

Certain observations follow from this table. The mean or effective QE is close to the peak QE for all scintillators emitting UV and blue light. This is because the

emission curves lie comfortably within the bialkali QE curve. The advantage of a green sensitive Rb bialkali is evident for ATP, BGO, CdWO₄, and CsI(Tl). An S20 photocathode is barely justifiable for ATP and CsI(Tl), taking note of the cost and high dark counts.

Matching the emission spectrum to that of the photocathode response does not necessarily lead to the highest fidelity: the signal-to-background ratio must also be considered. Thirty mm diameter PMTs, for example, generate dark counts in the following ratios: K bialkali:Rb bialkali:S20=50:200:2000 counts/s and this should be noted when selecting a PMT, particularly for low-light-level measurements. For example, the figure of merit for detectivity, universally adopted in low-light-level tritium detection, is E^2/B , where E is the signal count rate, and B the background—predominantly single photoelectron (see Chapter 4). However, by contrast in scintillation spectroscopy with sodium iodide, background counts contributed by the PMT are generally too low to affect energy resolution.

2.5.3 Solar blind photocathodes

The solar blind range of PMTs is characterized by a VUV-transmitting window and photocathodes suitable for the detection of radiation at VUV and UV wavelengths. The available window materials include lithium fluoride and magnesium fluoride, which are transmissive from 105 nm and 115 nm, respectively. PMTs with these windows are difficult to manufacture, primarily because of the mismatch in expansion coefficient between the window and the material of the envelope. Customers sometimes find to their dismay that the window has ‘popped off’ their rare PMT, despite careful storage and handling. In addition, LiF is subject to surface interaction with moisture, and sometimes it suffers from the inclusion of colour centres. MgF₂, on the other hand, is birefringent, with its c-axis lying in the window plane. It is therefore not surprising that manufacturers no longer offer LiF variants as catalogue items. Applications for these PMTs are principally in earth atmospheric physics, rocket plumes, and solar physics. Two photocathodes in this group, Rb₂Te and Cs₂Te, have similar spectral sensitivity covering 115 to 350 nm, while a CsI photocathode responds to wavelengths between 115 and 200 nm. Spectral response curves are shown in Fig. 2.20.

Consistent with their description, these photocathodes should not respond to visible radiation, and hence the expression ‘solar blind’. One measure of the quality of such photocathodes is the ratio of the peak QE to the QE at a stated wavelength in the visible, or near visible; optimal setting of this ratio is a trade-off between the demands of the application and the level of the customer’s funding. For Cs₂Te and Rb₂Te, this ratio is in the region of 10³ to 10⁴ (Charman 1969). CsI has a similar peak QE with a cut-off ratio of up to 10¹⁰. A desirable feature of the solar blind PMTs is their ability to detect the Lyman- α line at 121.6 nm, and specifications sometimes refer to this wavelength. PMTs with quartz windows, incorporating the three photocathodes, are available at lower cost, but transmissivity of the window only starts at 150 nm.

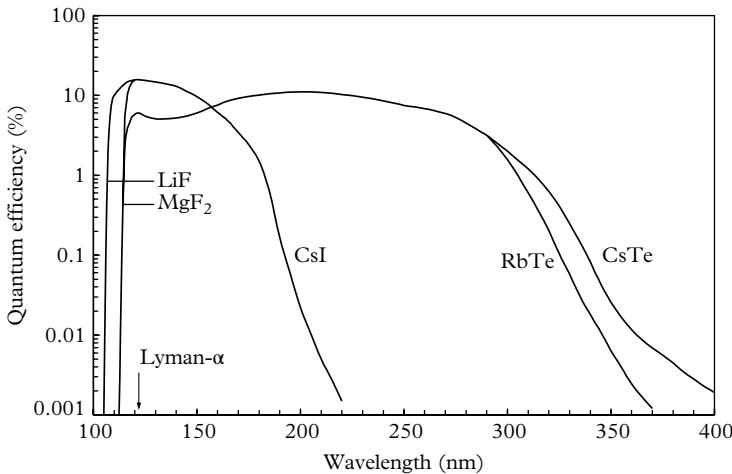


Fig. 2.20. Variation of QE with wavelength for the solar-blind range of PMTs. Note the approaching shoulder near the high wavelength cut-off for all photocathode types.

Canfield (1998) reviewed radiometric standards for wavelengths between 5 and 120 nm, which are outside the range of PMTs. This useful article covers the application of three detector types: photoemissive photodiodes consisting of a windowed device and an anode within an evacuated structure; silicon photovoltaics; and GaP/GaAsP are the preferred semiconductor devices, together with avalanche devices. It should be noted that QE calibrations performed by standards laboratories carry considerable uncertainty—of the order of 5% at wavelengths below 200 nm.

The gain of a PMT with a CsI photocathode is an order of magnitude lower than a bialkali PMT with the same number of dynodes; the SER is poor because of low d₁ gain. The reason for this lies in the manufacturing process for end window PMTs, in which the photocathode is laid down by evaporation from a small CsI bead, without disassociation. This process does not produce free Cs, and consequently the gain of BeO dynodes within the PMT is not enhanced by photocathode processing. However, Cs₂Te and Rb₂Te variants, with the same multiplier, have higher overall gain because Cs is admitted as a free element, which subsequently reacts with the BeO dynode surfaces, thereby increasing their secondary emission coefficient. Such a multiplier has a gain–voltage characteristic similar to those of a Cs-activated BeO multiplier. Solar blind PMTs are available in end window and side window constructions. A PMT with VUV and visible responses can be created by coating the outer surface of the window of a standard bialkali with wavelength shifter. As discussed in 3.13.6, incident light is absorbed by the coating and readmitted at a longer wavelength, matching the peak photosensitivity of the PMT. In practice, this type of light transformation can be realized with many different types of end window PMTs.

2.5.4 Infrared photocathodes

Until the introduction of III–V photocathodes, based on InGaAs, applications involving the detection of light in the region of one micron in wavelength were barely satisfied with an S1 (AgOCs) photocathode, while an S20 (Na₂KSbCs) photocathode has little sensitivity at these wavelengths. The reason for the poor response is the transparency of the photocathode to infrared light, as discussed in 2.4.6. The sensitivity curve for an S20 can be shifted to longer wavelengths by making a thicker photocathode; this increases the response where desired at the expense of blue sensitivity. Such photocathodes are sometimes referred to as S25, but response at 1 μm is still poor, even with the adoption of optical enhancement techniques. An S1 photocathode, one of the earliest types, has a QE of ~0.1% at 1000 nm, and one-tenth of this figure at 1100 nm. S1 PMTs have a reputation for poor stability and exceptionally high dark current, which is some three orders of magnitude higher than that from an S20 operated at the same gain.

The III–V photocathodes based on GaAs(Cs) provide ~1% QE at 900 nm, with a rapidly falling response beyond this wavelength. The introduction of InGaAs(Cs) photocathodes provided an alternative to the S1 with the attainment of 0.1% QE at 1040 nm. The recently introduced InP/InGaAsP and InP/InGaAs photocathodes have QEs of 0.5% at 1300 nm and 0.3% at 1500 nm, respectively (see Table 2.4). Note the lower two entries in Table 2.2(b) may be damaged by drawing in excess of 1 μA anode current. In addition, the PMTs must be cooled to at least –80 °C before switch-on. It is conventional to plot response curves as in Fig. 2.21, where the ordinate is in units of millamps per watt, using (2.5) for the conversion from QE to millamps per watt.

The millamps per watt scale highlights the sudden drop in sensitivity as cut-off wavelengths are approached. The essentially flat response of a GaAs(Cs) PMT, when displayed on a radiant sensitivity scale, is particularly desirable. The example

Table 2.4 A selection of PMTs for use in the red and near infrared. All photocathodes, except for the S20 and S1 types, are reflective and consist of a chip of semiconductor mounted within a 51 mm diameter PMT.

Type	Photocathode	Dimensions	λ_0 (nm)	Dark counts/s
9863/100, ETL	S20	2.5 mm dia.	1% (800 nm)	40
R316–02, HPK	S1	25 mm dia.	0.25% (900 nm)	2×10^7
R943–02, HPK	GaAsP	$10 \times 10 \text{ mm}^2$	1% (900 nm)	5×10^5
R3310–02, HPK	InGaAs	$10 \times 10 \text{ mm}^2$	0.3% (1000 nm)	6×10^6
R5509–43, HPK	InP/InGaAsP	$3 \times 8 \text{ mm}^2$	0.5% (1300 nm)	1.6×10^4
R5509–73, HPK	InP/InGaAs	$3 \times 8 \text{ mm}^2$	0.3% (1500 nm)	1.6×10^5

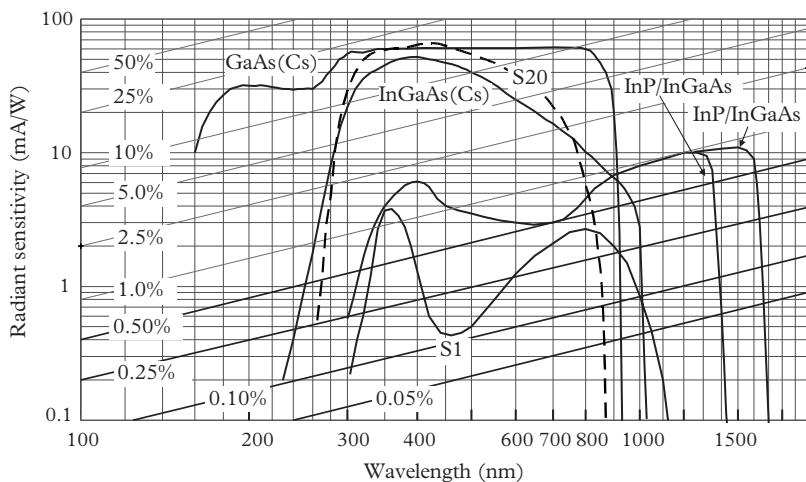


Fig. 2.21. Infrared photocathode response curves. The diagonal lines represent QE.

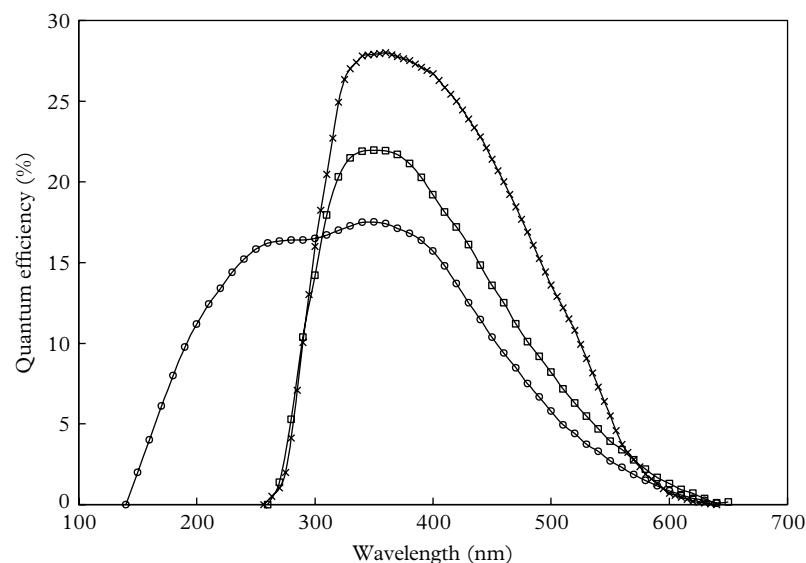


Fig. 2.22. Spectral response of high-temperature bialkali photocathodes; $-x-$, standard low-temperature bialkali; $-\square-$, high-temperature bialkali with borosilicate window; $-o-$, high-temperature bialkali with sapphire window.

shown for GaAs(Cs) refers to a PMT with a silica window extending the sensitivity in the UV to 160 nm. A similar extension applies to all other photocathode types, except for the S1.

2.5.5 High-temperature photocathodes

PMTs with bialkali photocathodes can be operated at temperatures approaching 100°C, although with a reduced lifetime. Beyond this temperature, it is claimed that caesium dissociates from the photocathode, thereby lowering its sensitivity. A high-temperature bialkali photocathode, Na₂KSb, does not contain caesium and consequently it may be operated at temperatures approaching 200°C. The application of this type of PMT in oil well logging is discussed in 12.2.2. We note that the spectral response is similar to that of the standard bialkali (K₂SbCs), but depressed by about 30%, as shown in Fig. 2.22. The window material is commonly glass, but the ultra-rugged high-temperature metal-ceramic version includes a sapphire window for robustness.

2.6 Photocathode geometries

2.6.1 Range of end window PMTs

End window photocathodes consist of a thin, semi-transparent film of semiconductor, deposited on the inside of the entrance window. These are called transmission types, because the incoming radiation and the photoelectrons travel in the same direction. A selection of geometries, (a) to (l), is given in Fig. 2.23. In the majority of types, including MCPPMTs, the active area is of circular geometry.

- (a) These PMTs incorporate a plano-planar window with photocathode diameters ranging from 9.7 mm (3/8") to 133 mm (5").
- (b) Also in this group are those of reduced active area, achieved by electrostatic defocusing. Envelope diameters are either 28mm (1 1/8") or 52mm (2"), with active photocathode diameters of 2.5 mm or 9 mm, respectively. Advantage is taken of a much reduced dark current in applications where the light is tightly focussed or collimated.
- (c) Plano-concave windows provide better timing than the plano-planar equivalents.
- (d) 2π PMTs collect light over an increased solid angle, by withdrawing aluminizing to part way down the envelope. PMTs are available in an envelope diameter of 28 mm.
- (e) All circular geometry options are available with a thin meniscus lens in place of a plano-planar window. The lens is only ½ mm thick in a 52 mm diameter

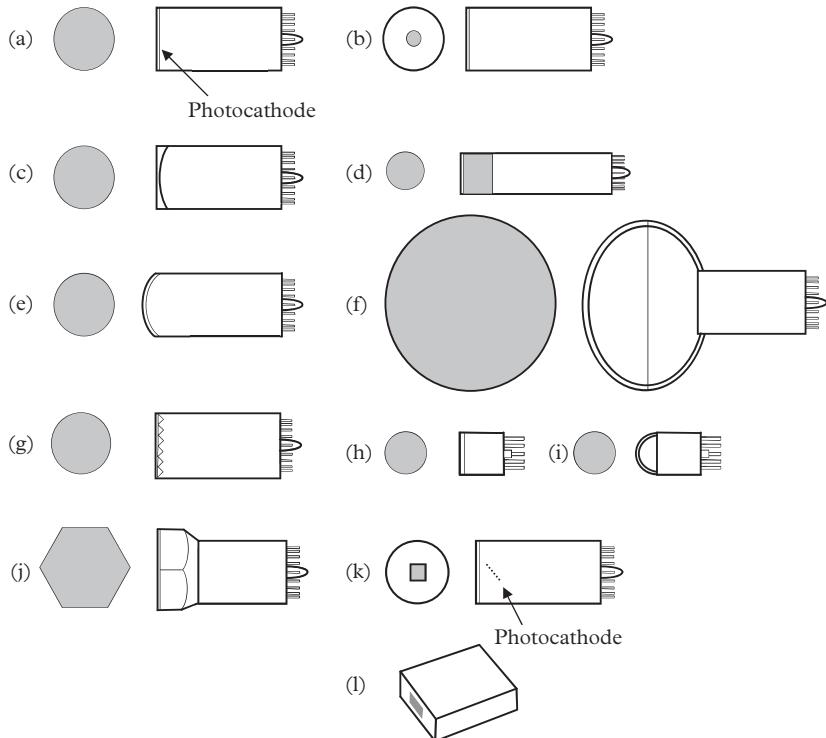


Fig. 2.23. Examples of end window photocathode geometries.

PMT, compared with the standard of 2mm for a plano-planar window. A thin window reduces the size of Cerenkov pulses, derived from the natural background or from signal. Radioactivity is also reduced, and in both cases the benefit scales with the inverse of the thickness. Meniscus lens variants are available in the larger-diameter PMTs, with a thickness of 2 to 3 mm.

- (f) Hemispherical PMTs are an extreme variant of a meniscus window, finding application in fundamental scientific research, usually immersed in a large volume tank of liquid, such as water, scintillator, or liquid argon. A 3 mm hemispherical window is particularly robust and able to withstand two atmospheres of pressure. The smallest available diameter is 25mm (1"), used in radiation monitors, and the largest is 500mm (20"), chosen for massive neutrino detectors.
- (g) A plano-planar window with an embossed pattern of prisms gives enhanced performance for red and infrared collimated light in the manner discussed in 3.13.4. They are available in diameters of 28mm and 52mm.
- (h) Metal package PMTs offered by Hamamatsu, the R7400 series, are small in diameter and in length. These function in the same way as the all

glass end window PMTs except that the envelope, of 16mm in diameter, is metal.

- (i) Multi-anode variants in the R5900 series are also available with a hemispherical or rectangular window.
- (j) Square, rectangular, and hexagonal front-end geometries include a planoplanar window devised for applications where close packing allows the maximum coverage of large light-emitting areas, such as a gamma camera.
- (k) The GaAsP (type R943-02) and the InGaAs (R3310) photocathodes are in chip form (10mm × 10mm) and enclosed within an end window envelope. These two PMTs do not conform to the definition adopted in this section, but they are designated end window for convenience. This also applies to the next entry.
- (l) The Hamamatsu H12400-00-01 assembly contains the smallest PMT currently available. The microPMT® has dimensions of $13 \times 10 \times 2\text{mm}^3$ with a $3 \times 1\text{mm}^2$ active area photocathode.

This list is not exhaustive, and users should check the latest product on offer, by either looking in the manufacturers' catalogues or consulting the internet.

2.6.2 Range of side window PMTs

Side window photocathodes are deposited on a rectangular metal substrate located within a glass envelope in the shape of a test tube (Fig. 2.24). Photoelectrons are emitted from the illuminated side of the photocathode strip, and the cathode is referred to as reflective.

The original side window PMT was the first commercially available PMT: the RCA 931A was launched in the 1930s for reading the sound stripe of talking movies. The design of current side window PMTs has changed little over the years; the notable difference, however, lies in performance. In the original designs, the photocathode consisted of a rectangular strip of dimensions $8 \times 24\text{mm}$, which remains the standard size. In the 1970s Hamamatsu offered the R928; this PMT

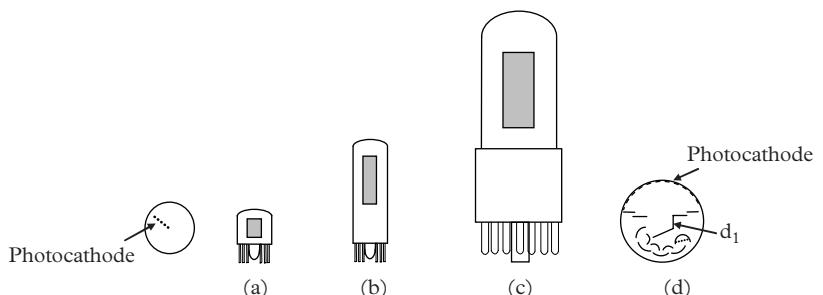


Fig. 2.24. Examples of side window photocathode geometries.

has an S20 photocathode of ultra-high performance, which no other PMT manufacturer has to date been able to match. Side window PMTs are available in various photocathode sizes from the compact 3mm × 4mm to the original 8mm × 24mm format. All photocathode types—VUV, UV, visible, and III–V—are available in the side window geometry:

- (a) Referring to Fig. 2.24, the smallest side window PMT is the R3810, with an opaque 3mm × 4mm photocathode in an envelope of 17mm seated height.
- (b) The R6350 has a seated height of 24mm and a photocathode of 4mm × 13mm.
- (c) The original side window design has a photocathode of dimensions 8mm × 24mm.
- (d) Another side window envelope design has a transmission photocathode deposited on the glass of the bulb; for example, the R5108 has an S1 photocathode in this configuration.

2.6.3 Window materials

The radiant sensitivity of a PMT depends on the wavelength of light and on the transparency of the window at this wavelength (Fig. 2.25). All materials suitable for use in PMTs have a lower cut-off wavelength below which transmission falls rapidly. Above 350 nm, all materials, with the exception of sapphire, transmit light without discernible loss up to a wavelength of about 2000 nm. However, transmission properties are quite different at UV wavelengths, as is evident in Fig. 2.25. Borosilicate glass, for example, approaches full transmission above 350 nm, compared with ~200 nm for fused silica. Historically referred to as quartz, a naturally occurring crystal, indicated by the suffix Q, is still used in PMT type numbers to designate fused or synthetic silica. The data in Fig. 2.25 include Fresnel reflection losses (see Fig. 2.8), although graphical presentations in the scientific literature are sometimes normalized. Window thickness affects transmission and hence QE, particularly at UV wavelengths.

It is worth mentioning at this point that most glasses are attacked by pure water. This is of concern for the long-term deployment of large-area hemispherical PMTs in massive water Cherenkov tanks. Special glasses such as Schott 8245, used for SNO, and ET Enterprises B53, selected for the Borexino experiment in Gran Sasso, exhibit low levels of radioactivity and are resistant to water attack. Such glasses do not include potassium compounds and are consequently almost free of the long-lived isotope ^{40}K . However, the omission of this element makes the glass more difficult to work, leading to visual impairment in the form of ‘cords’, ‘striations’, and sometimes ‘bubbles’. Such defects are only important in PMTs with imaging capability.

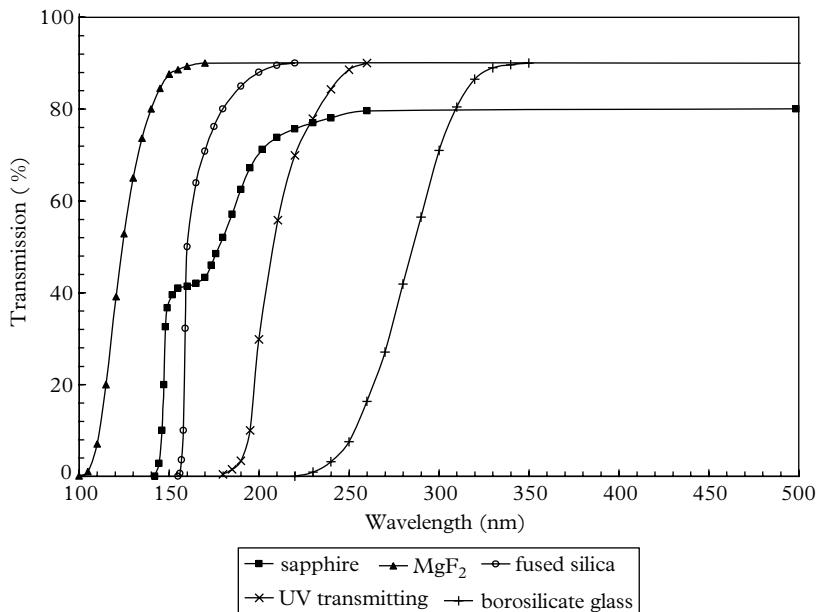


Fig. 2.25. Transmission curves for window materials of 2 mm thickness, including reflection losses. Windows range in thickness from 1 to 10 mm, depending on the diameter of the PMT.

2.7 Photocathode resistivity

An end window photocathode is limited in the current that can be drawn because of its resistivity. There are two considerations: the materials used for fabrication, and the thickness of the photocathode layer. Solar blind photocathodes, such as CsI, Cs₂Te, and Rb₂Te are essentially insulators and worthless as photodetectors, unless laid down on a conducting substrate. It must be sufficiently thin to ensure that 80 to 90% of incident light is transmitted to the photocathode and yet sufficiently thick to provide adequate conductivity. Substrates are generally evaporated on to the window prior to photoactivation, using metals such as tungsten, chromium, or platinum. A photocathode is deemed resistive should it show a dependency on the $\nu(k-d_1)$ voltage; the lower curve in Fig. 2.26 is clearly non-linear.

Resistivity, R_{\square} , is measured in ohms per square, which quantifies the resistance of a surface film between conductors on the opposite side of a square layer (note that R_{\square} is independent of the units chosen for the side length). The resistance per square for various semi-transparent photocathodes, as a function of temperature, is given in Fig. 2.27, but there is considerable variation even among PMTs of the same type. The effect of a platinum substrate, in reducing the resistivity of a

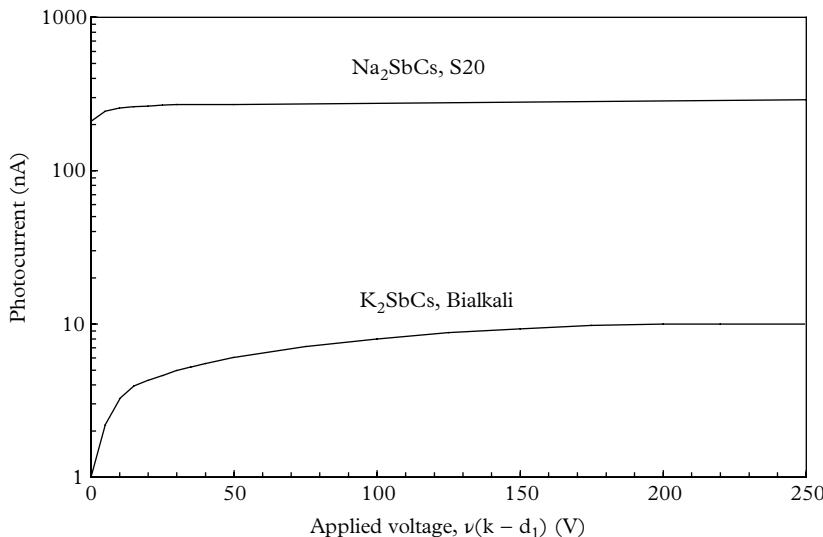


Fig. 2.26. Measured photocurrent for a fixed light input, as a function of applied voltage, $v(k - d_1)$. The effect on collection of a resistive K₂SbCs photocathode compared with an S20 type is significant even at room temperature. Curves such as these, for 50 mm diameter PMTs, are the basis for the recommendation of limiting photocurrent to $<< 1$ nA at room temperature.

bialkali photocathode, is clear from Fig. 2.27 by comparing the curves marked K₂SbCs and (Pt)K₂SbCs.

The use of a standard bialkali PMT, that is one without a substrate, is limited to applications in which the operating temperature remains above about -100 °C. Similarly, there is a temperature limitation, although less severe, on the use of the Rb-based photocathodes. The design of an input electron-optical system between the cathode and the first dynode assumes that the photocathode is an equipotential surface. Resistive photocathodes fail to meet this requirement, once appreciable photocurrent is drawn. A calculation can be made to estimate the voltage drop between the periphery and the centre of an end window PMT photocathode. According to Piétri (1994), ΔV is given by

$$\Delta V = R_{\square} I_k / 4\pi. \quad (2.21)$$

For example, if $R_{\square} = 10^{10}$ Ω/square with a photocurrent of 1 nA, ΔV turns out to be of the order of 0.8 V and close to the typical photoelectron emission energy of ~ 1 eV. Consequently, a photocathode will only present an equipotential surface under low-light-level conditions, corresponding to a photocurrent, I_k , of less than 0.1 nA in this example. The paths of photoelectrons released from the edge of the photocathode are particularly sensitive to any change in electron-optical conditions. Figure 2.26 illustrates the effect on collection in two different PMTs,

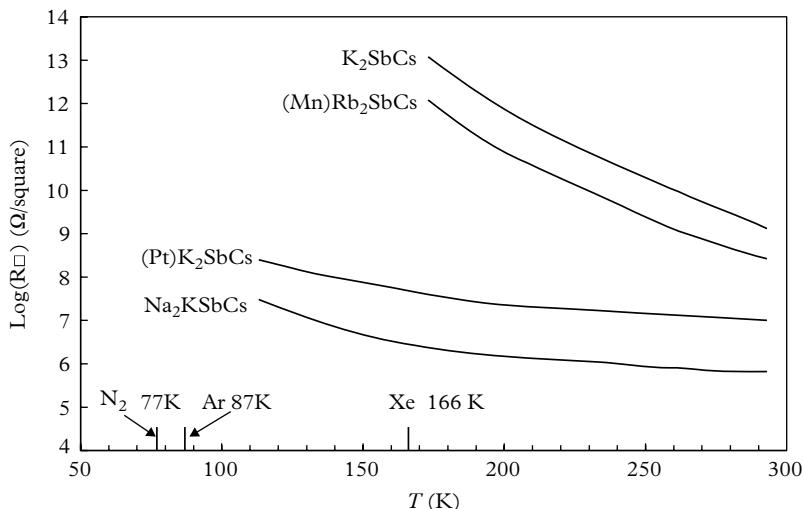


Fig. 2.27. Measurements made using vacuum diode photocells in which photocathode material is deposited between two parallel contacts. The elements within brackets indicate the inclusion of a substrate. A J Cormack (private communication).

operated as photodiodes, highlighting the dissimilarity in behaviour between a bialkali and an S20 photocathode.

The actual shape of these curves is particularly sensitive to the area of illumination and marginally dependent on wavelength. The characteristics shown in Fig. 2.26 refer to uniform illumination over the entire photocathode area of a 50 mm diameter PMT. Note there is always partial collection from energetic photoelectrons emitted at favourable angles, even in the absence of an electric field. Resistivity effects are mitigated in PMTs for which the electron-optical design provides a relatively high extraction field gradient at the photocathode (e.g. XP2020).

We now examine the implications of high resistivity in the applications of a bialkali PMT—the type particularly sensitive to these considerations. The limited dynamic range in photocurrent is unlikely to restrict performance at room temperature, once it is appreciated that a photocurrent of the order of 0.1 nA corresponds to an intense light signal, easily visible even if uncollimated. PMTs are designed for use at gains in excess of 10^4 , and operation at a gain lower than 10^4 should prompt the question as to whether a PMT is the correct choice of detector. The corresponding anode current is 1 μ A in this example, and the PMT would fatigue for reasons unrelated to resistivity. On the other hand, R_{\square} for an S20 photocathode is four orders of magnitude lower than for a bialkali and thus poses no restriction on photoelectron collection at low temperature.

These concerns refer to steady and slowly varying photocathode illumination, but relatively intense pulsed currents can be handled with acceptable linearity, provided the mean rate is sufficiently low and consistent with these considerations.

For example, a reasonable assumption of the stray capacitance between a photocathode and ground of 1 pF/cm^{-2} leads to $\Delta V \sim 16 \text{ mV}$ for an intense flash of 10^5 pe . Clearly, this small drop in potential has no effect on collection, although the recharge time may be significant, should a resistor be placed in series with the photocathode pin; this is sometimes adopted for overload protection. The foregoing discussion is relevant to end window, semi-transparent photocathodes only. The resistivity of reflection-mode photocathodes, such as those in side window PMTs, is very low because these are laid down on a substrate of metal alloy. The resistivity of some common photocathodes is shown in Fig. 2.27.

2.8 QE temperature dependence

2.8.1 Laboratory and commercial applications, –30 °C to +60 °C

The effect of temperature on the radiant sensitivity of PMTs engaged the attention of scientists in the late 1950s working in the field of gamma ray spectroscopy, for example. The paper by Murray and Manning (1960) summarizes the essentials of the subject, with conclusions largely in agreement with present understanding. They studied Dumont and RCA PMTs with S11 and S20 photocathodes, for at that time the bialkali photocathode was unknown. The majority of PMT applications fall within $\pm 50 \text{ }^\circ\text{C}$ of room temperature, with corresponding temperature coefficients for QE given in Fig. 2.28. It is immediately apparent that $\alpha_{\text{QE}}(\lambda)$ is small and therefore not easily measured, except near the long wavelength cut-off, λ_0 . At such wavelengths there is a significant loss in QE with temperature. The long wavelength limit is usually taken as the wavelength at which the QE has fallen to 1 % of its peak value. Figure 2.28 warns of a trap that some users fall into through excessive cooling of a PMT to enhance the signal-to-background parameter; this is common in applications involving red and infrared wavelengths. Cooling only slightly enhances the QE at short wavelengths but the reverse applies at long wavelengths. The guiding principle is, therefore, cool no more than necessary. A further historical recommendation is to cool uniformly, for it is reported that cooling solely the window results in caesium migration and permanent loss of photocathode sensitivity.

The reasons for limiting the level of cooling become clear on examining the changes in the background pulse height spectrum. Referring to Fig. 2.29, for a bialkali photocathode, it is apparent that regions A and B show a reduction in single photoelectron dark counts by a factor of about 3, while the counts in the multi-photoelectron bands C and D are independent of temperature. An S20 photocathode shows a decrease in dark counts in regions A and B of up to three orders of magnitude, but there is little additional benefit in cooling below 0 °C.

66 Photocathodes

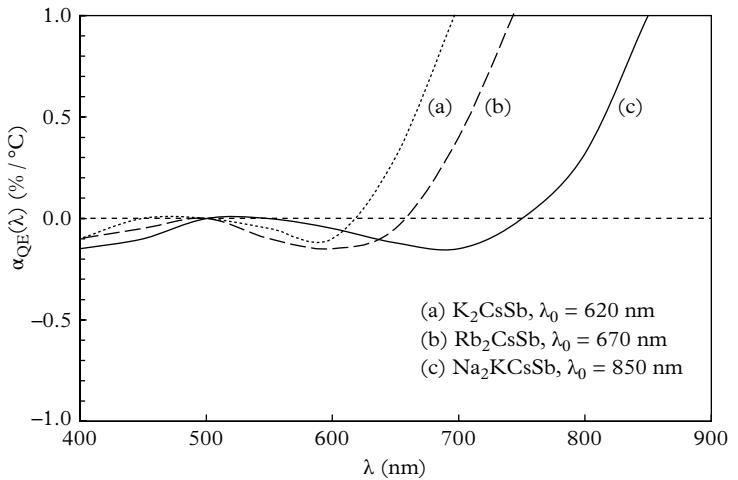


Fig. 2.28. Temperature coefficient for three main photocathode types. A loss in QE corresponds to a positive $\alpha_{QE}(\lambda)$. Cut-off values, λ_0 , are indicated in the figure.

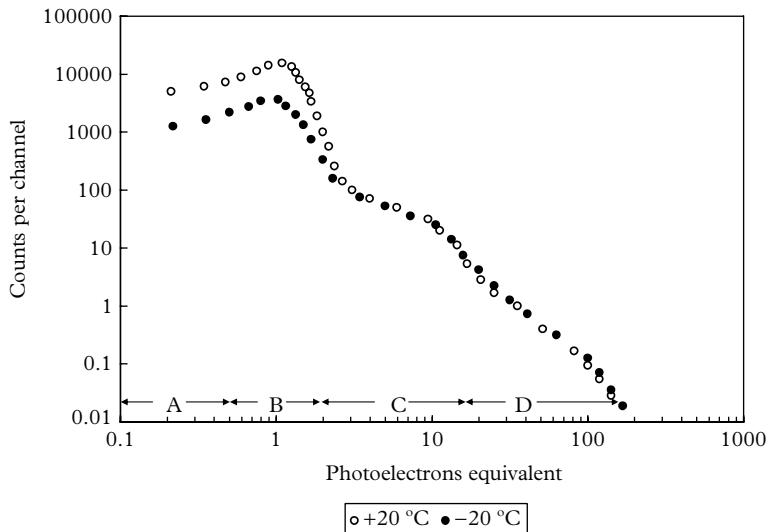


Fig. 2.29. The effect of cooling on the background spectrum of a bialkali PMT. Contributions in regions A and B are defined and discussed in 6.6 and are clearly temperature sensitive.

In principle, photocathode temperature coefficients should be measured in a photodiode configuration—the same method adopted for QE. However, this implies low-level photocurrents to avoid the effects of resistivity. Unavoidable leakage currents pose additional complications. It is therefore readily understood

why photocathode temperature sensitivity is usually determined with the multiplier active, without realizing that the latter is also temperature sensitive. The method described by Singh and Wright (1987), based on photon counting, isolates these contributions: the count rate is a measure of the change in QE and the shift in the peak, or mean pulse height of the SER, represents gain change. This effect is small and hence difficult to determine accurately, with α_{g} of the order of $-0.2\% \text{ K}^{-1}$ over the temperature range $+30^\circ\text{C}$ to -20°C , with gain increasing on cooling.

2.8.2 Ultra-low temperature performance, -273°C to -30°C

QE increases for blue light, is stable for green light, and decreases monotonically for red light, over the temperature range stated in 2.8.1 and illustrated in Fig. 2.28. There is a cut-off in QE at ~ 160 K for standard bialkali photocathodes (K_2CsSb), which is attributed to photocathode resistivity, discussed in 2.8.1. Stable PMT performance extends to at least 77 K (liquid nitrogen) by using a specially fabricated PMT that includes a thin platinum substrate for the photocathode (designated (Pt) K_2CsSb). The ultra-low temperature 9357 FLA hemispherical PMT shows a small increase in the QE for blue light, consistent with $\alpha_{\text{QE}}(\lambda) = -0.15\% \text{ K}^{-1}$, down to ~ 200 K; thereafter, QE decreases linearly at a rate of $+0.2\% \text{ K}^{-1}$ to liquid nitrogen temperatures. Studies by Ankowski *et al.* (2006), covering the performance of 54 ETL 9357 FLA hemispherical PMTs, are particularly relevant. These hemispherical PMTs, chosen for the experiment ICARUS, show that all critical PMT parameters—cathode sensitivity, anode pulse shape, SER, linearity, dark counts, and afterpulses—are not significantly affected by operation at 77 K.

McMillan and Reid (1989) measured 22 standard ETL 9821B bialkali PMTs operated at 193 K. A pulsed LED provided nominally 15 pe/pulse. Figure 2.30 indicates that performance at this temperature, at the chosen light levels, is comfortably within the range of linear performance. These authors observe a hysteresis effect between cooling and recovery cycles, whereby the gain after cycling is different from the initial gain. This is a 10% effect but the gain recovers after a few days. An alternative to the use of a conducting semi-transparent substrate is to include thin aluminium fingers evaporated on to the inside of the window and disposed radially. The fingers are ~ 0.5 mm wide and terminate short of the central region of the photocathode. They obscure approximately 2% of the photocathode area and in this sense they are more effective than a continuous thin substrate. The ZEPLIN III dark matter experiment described by Araújo *et al.* (2004) operated at 166 K using a modified ETL 9829B (D730), which includes aluminium fingers. It is curious that the fingers actually produced an effect opposite to expectation, although the authors observed a faster recovery from saturation.

The distinctive feature in Fig. 2.30 is the sudden drop in sensitivity at temperatures below 160 K, as the influence of resistivity becomes significant. At temperatures above 160 K, conductivity increases and the three curves follow the well-established

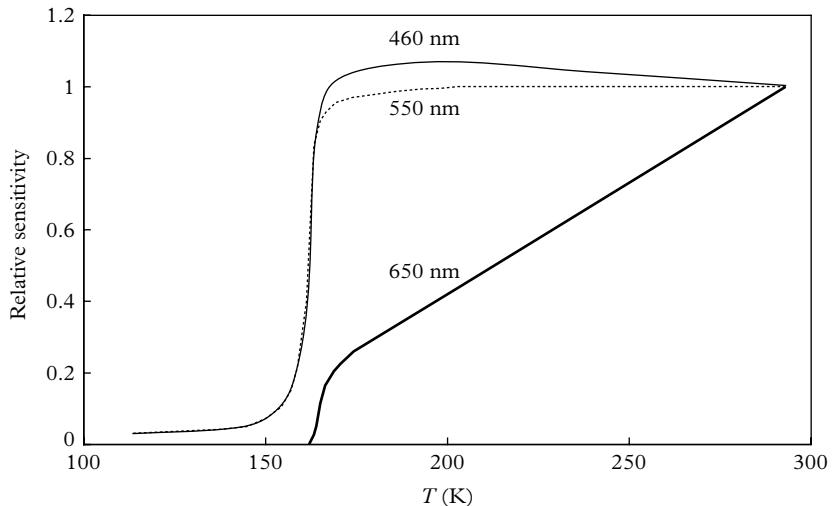


Fig. 2.30. The relative photocathode sensitivity of an EMI 9750B bialkali, as a function of temperature, for three wavelengths. Pulsed LEDs were used to derive these curves. Such graphs are referred to as ‘saturation curves’.

sensitivity relationship with temperature coefficients of $\alpha_{QE}(\lambda) = -0.15\% \text{ K}^{-1}$ for blue light, $+0.6\% \text{ K}^{-1}$ for red, and essentially 0 for green light. The precise temperature at which the transition occurs depends on the intensity of the light and whether it is steady or pulsed. Araújo *et al.* (1997) show this very nicely with a green LED operated at 10 kHz and at 100 Hz: the transition point moves from -100°C to -150°C at these two rates. This rate effect is also evident in the measurements of Ichige *et al.* (1993) over the range 1 to 100 kHz. The measurements reported by Araújo *et al.* (1997) show unwelcome fine structure superimposed on the curves of the type in Fig. 2.30. This applies to the products from three manufacturers. The authors attribute these sudden changes in sensitivity to intermittent photocathode contact, which is a likely explanation since the contact is a mechanical one in all end window PMTs. No such anomalous behaviour is observed in a batch of 54 ETL 9357 hemispherical PMTs operated in liquid nitrogen, and reported on by Ankowski *et al.* (2006). The successful deployment of significant numbers of PMTs in a variety of massive detectors indicates that manufacturers have clearly risen to the challenge by producing reliable products suitable for use at cryogenic temperatures.

2.8.3 High-temperature operation, $+60^\circ\text{C}$ to $+200^\circ\text{C}$

Standard bialkali PMTs perform satisfactorily up to temperatures of about 100°C , although with a shortened life. Oil well logging applications, for example, requires operation up to 200°C and demand a special PMT for this purpose.

The high-temperature bialkali, briefly discussed in 2.5.5., is uniquely suitable, but there is a considerable reduction in QE in the region of 420 nm, the predominant wavelengths of interest. The temperature dependence of $-0.15\% \text{ K}^{-1}$ generally applies up to about 100°C, at which temperature the QE is ~85% of that at room temperature. Above 100°C, $\alpha_{QE}(\lambda)$ increases rapidly with temperature such that, at 175 °C, the QE is reduced to ~40% of its room temperature figure (Persyk *et al.* 1976).

2.9 Dark count temperature sensitivity

There are few reported measurements on the behaviour of dark counts at cryogenic temperatures. Araújo *et al.* (2004) noted a decrease of typically 25% between room temperature and 173 K. Had they made measurements below 100 K they would undoubtedly have observed an increase in dark counts as reported by Ankowski *et al.* (2006). In the investigations by the latter group, an increase in dark counts by a factor of about 3 at 77 K over that at room temperature is evident. These authors explain the effect in terms of ‘a decrease in lattice energy of the photocathode material at low temperature resulting in an increase in escape probability for electrons’. Meyer (2010) cooled a Hamamatsu R7725 PMT to 4 K, observing an exponential increase in dark counts between 250 K and 4 K. This author reports that the dark counts appear in bursts and are consequently highly correlated.

In summary, making measurements at cryogenic temperatures is costly in the use of liquid nitrogen, exacerbated in the time required to reach stable PMT performance following a change in operating temperature. Authors comment on the difficulties and uncertainties in the time required to achieve stable performance—ranging from hours to days. It is interesting to note that manufacturers still regard -30 °C to +60 °C as the safe operating range for PMTs, despite evidence from dark matter and astrophysics experiments that a much lower temperature of operation is attainable.

2.10 Critical photocathode properties

2.10.1 Photocathode uniformity

The concept of photocathode uniformity of response is not straightforward. It is necessary to distinguish between the intrinsic uniformity in QE of the photocathode itself, and the signal produced at the anode by scanning the photocathode with

70 Photocathodes

a spot of light. The user, naturally, is primarily concerned with the uniformity observed at the anode. Cathode and anode non-uniformity are different because of losses and non-uniformities caused by the electrostatic focussing elements located between the photocathode and the first dynode, and the asymmetry in the dynode multiplier itself. Intrinsic cathode uniformity is measured by connecting all electrodes, other than the photocathode, to the anode, with the PMT thus operating as a photodiode. Uniformity may be quantified by scanning a small spot of light across the photocathode; alternatively, a pinhole source may be moved manually over the plane of the window. The wavelength of the light used should be close to that of the intended application—especially important where the transmission of UV through a plano-concave window is concerned. This is because the increased thickness near the periphery reduces the light transmitted to the photocathode area as the edge is approached. The intensity of the spot must be chosen sufficiently low to avoid the effects of photocathode resistivity, which is covered in 2.7. Generally, the photoelectron response of the photocathode is uniform, although there may be a small degree of wavelength dependency and enhancement around the perimeter of the window. The latter originates in the window-envelope interface region, through the combination of an external rounded shoulder and aluminizing within the envelope. The effective photocathode diameter in end window PMTs is determined by a combination of the inner diameter of the glass envelope and the electron-optical design. Ideally, the QE should be uniform within the manufacturer's specified effective photocathode diameter. The effective dimensions are always smaller than the outside diameter of the PMT, and in scintillation spectroscopy this is the reason for recommending a PMT of diameter greater than the crystal to which it is to be coupled. An example where cathode uniformity is always critical is that of a gamma camera. This medical equipment uses a set of PMTs numbering up to 100, and usually of the hexagonal design, of the type shown in Fig. 2.23(j). PMTs are mounted on a single, thin, large-area sodium iodide crystal and the position of every gamma captured by the scintillator is determined by an algorithm that assumes uniformity in the individual photocathodes. For detailed information on gamma cameras, the reader is recommended the short account in the PMT handbook issued by Philips (Piétri 1994, Section 7.1).

The scan shown in Fig. 2.31(a) illustrates the edge effect in the form of ‘ears’ often seen in large-diameter PMTs. The remaining non-uniformities originate in the multiplier or caused by uneven evaporation of antimony in the course of photocathode activation. The second scan, shown in Fig. 2.31(b), is typical of a PMT with Venetian blind dynodes: the low response seen on the left of the diagram is caused by photoelectrons that bypass the first dynode through approaching the plane of the slats near tangentially. The corresponding signal at the anode is reduced on average by the first stage gain for such trajectories (this is further discussed in 5.6.1). There is a similar effect in large-diameter linear focussed PMTs where photoelectrons launched from a particular area of the photocathode approach the plane of the first dynode at a glancing angle, landing on the back of the third dynode. RCA remedied this problem in the 1970s by introducing the T-cup first dynode.

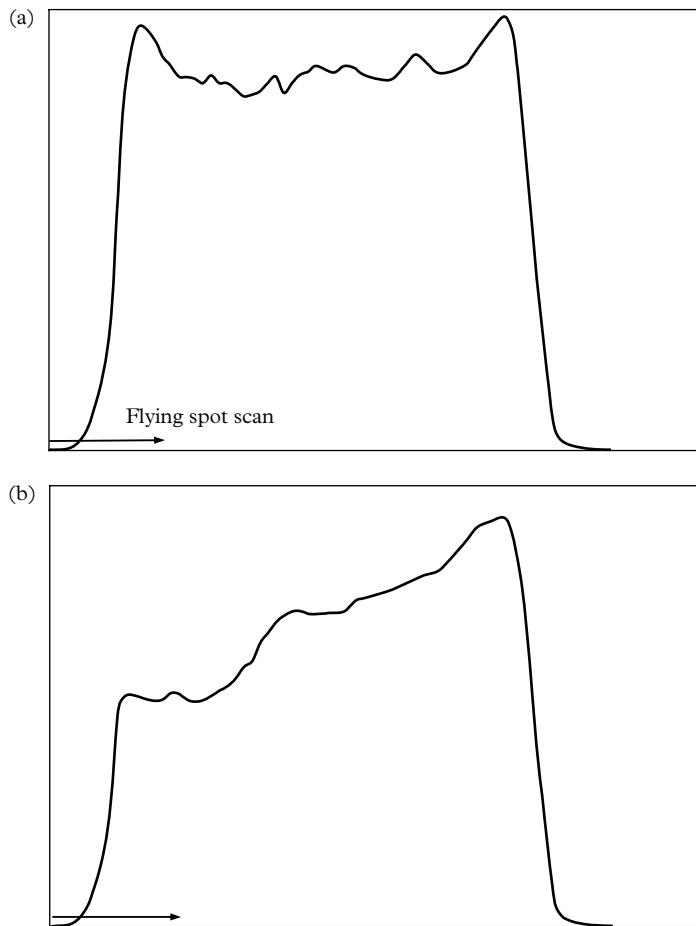


Fig. 2.31. Photocathode uniformity patterns measured at the anode of a Venetian blind PMT. The flying spot scan is across a diameter: (a) along the direction of the dynode slats, showing the ‘ears’ around the periphery as a result of unintentional optical enhancement; (b) across the direction of the dynode slats, showing the fall-off in uniformity because of the asymmetry of the first dynode.

As the name suggests this is a dynode of considerable area designed for high collection efficiency, but at the expense of speed—a parameter of no importance in the intended applications. The other consideration is one of finding the optimal compromise between speed of response and collection efficiency. An example of this is the long-established XP2020 with a collection efficiency of about 50%. This parameter has been sacrificed to achieve a low transit time jitter of only 250 ps. The XP2020 PMT was superseded by the XP2020 UR (Ultra Rapide) in the 1990s with improved collection together with jitter of only 150 ps (see 8.2.1).

2.10.2 Ionizing radiation

The consequences of high-energy radiation, such as gamma rays, neutrons, X-rays, electrons, and protons, on the transmittance of glass components are of prime importance in space exploration, military, scientific, and certain commercial applications. Transmittance refers to the transmission of light through a medium, as a function of wavelength. PMTs are deployed in space vehicles operating in an environment of energetic electrons and protons, trapped in the van Allan belt, for example, and there are regions in space where the high level of ambient radiation necessitates temporarily disabling the PMT HV. The ever-increasing radiation levels associated with the evolution of accelerator experiments pose similar problems, for example, from hadrons in the operation of the CMS electromagnetic calorimeter (Bell *et al.* 2001). The basic requirement in high-radiation environments is that of stable PMT performance over time. This prescribes that photocathode sensitivity and window transmittance must remain unaffected by irradiation. There is no evidence of permanent photocathode damage, although most of the optical glasses used in PMTs darken when exposed to excessive irradiation. UV-transmitting glasses have lower impurity levels than the more common borosilicate ones and they consequently show a smaller loss in transmission. Likewise, fused silica and magnesium fluoride windows suffer little change in transmission after a similar dose of radiation. Optical glasses, such as Schott 8245, suffer a shift of ~ 30 nm at the 50% transmission point, at ~ 350 nm after 100 Gy exposure (1 Gy = 100 rad), whereas fused silica shows no loss in transmission up to 10^6 Gy. The change in spectral transmission of a borosilicate glass which is widely used in PMT manufacture, Schott 8245, is shown in Fig. 2.32.

There is an extensive literature on this subject, a consequence of having to meet the demands of high-energy physics and space exploration. The work of Liepmann *et al.* (1992) and others contains considerable detail covering a range of materials. The redeeming feature in the loss of optical transmission is that the process is partly reversible, either with the passage of time, by heat treatment, or the application of UV radiation. The natural recovery in transmission, after a week from terminating a dose at 100 Gy, is shown in Fig. 2.32. Further examples may be found in the work of Liepmann *et al.* (1992), who have measured the recovery of a selection of glasses following treatment at 400°C. When annealing bialkali and S20 PMTs, the temperature should not exceed 100°C to avoid damaging the photocathode, but the process will obviously take longer at the lower temperature.

The effect on materials used for optical coupling between a PMT and a light guide or scintillator also needs consideration. Kobayashi *et al.* (1991) have made a thorough examination of the radiation resistance of such compounds. They investigated the transparency of both viscous fluids and adhesives over the range 200 to 600 nm. Thirteen samples were tested at exposures of up to 10^6 Gy, with thicknesses ranging from 4 to 10 μm . In general, thicker films are required to couple large-diameter PMT windows to allow for bowing under atmospheric pressure. Silicone elastomers, RTV, and silicone oils, the main compounds of

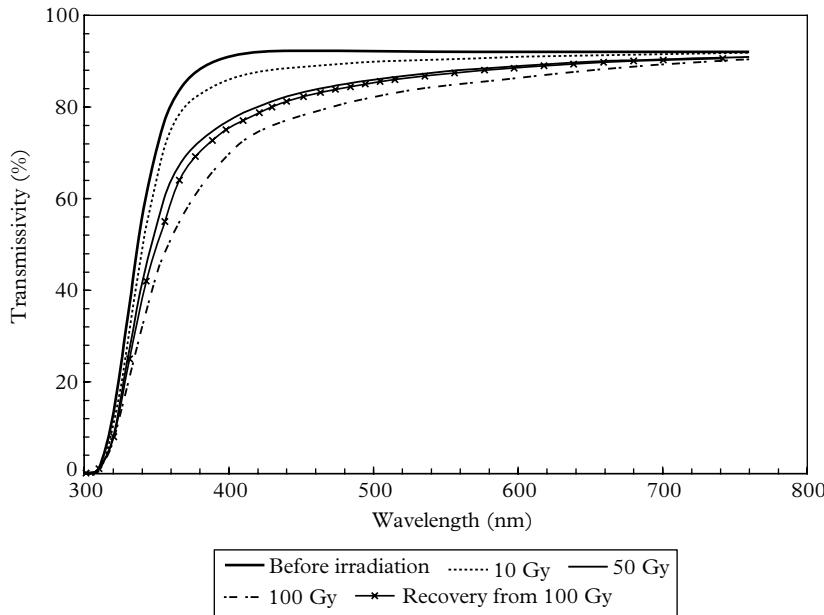


Fig. 2.32. The effect of radiation browning in a sample of borosilicate glass following exposure to various levels of 180 keV X-ray radiation.

interest, are barely impaired by doses below 10^5 to 10^6 Gy and, at such levels, window materials other than fused silica, would in any case be dark brown in colour. These authors relate an interesting observation concerning the hardening of silicone oils after exposure to 10^6 Gy of ^{60}Co , with no apparent recovery.

2.11 Summary of factors affecting photocathode response

Among the earliest measurements, concerning the dependence of QE on incident angle, is that of Hoenig and Cutler (1966). Based on S11 photocathodes, their results favoured parallel polarization, particularly at incident angles beyond 60° . Some 10 years later, measurements by Jones (1976) on S20 photocathodes were broadly in agreement with these authors, although not with their theoretical predictions. Lay (1997), continuing the studies of Moorhead and Tanner (1996) on the optical properties of bialkali photocathodes, concluded that ‘quantum efficiency and absorptivity, as a function of angle, are numerically the same to within a normalization constant’. On this basis, it is possible to compare theoretical

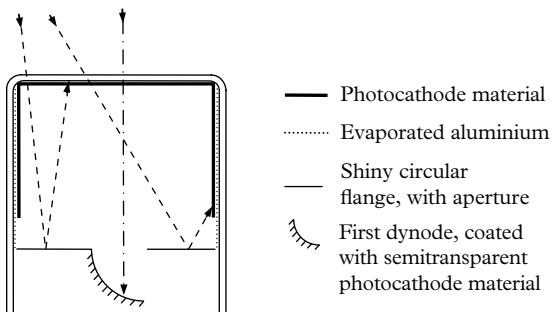


Fig. 2.33. Illustrating how light transmitted through the photocathode (---) is reflected on to photosensitive areas of the front end. The probability of reflection of the ray (- - -) by d_1 is small, but photoelectrons are generated from both the photocathode and the first dynode in this instance.

predictions with measured cathode sensitivities. Lay's results (his figures 10 and 11) also favour parallel polarization at glancing angles, consistent with the findings of these authors. Furthermore, the entire theoretical response curve, over all incident angles from 0° to 90° , is in agreement with measurement for both parallel and perpendicular polarizations, after allowing for the contribution from light that passes through the PMT window. Most end window PMT designs include a highly reflective internal support flange, in the form of a disc with a central opening, located in front of the first dynode. Some transmitted light is reflected by this flange and intercepted by the photocathode. A fraction of transmitted light interacts directly with the aluminized and photosensitive sidewall producing photoelectrons. Investigations with a flying spot show a dip of about 10% in response where the beam passes over the aperture; this effect is more marked with red light and gives an indication of the magnitude of this reflected component. Clearly, these contributions are dependent on the angle of incidence of the incoming light and on the transmittance of the photocathode at the assumed wavelength. Examples of these enhancement paths are shown in Fig. 2.33.

In general, enhancement of photosensitivity by these mechanisms is unwelcome, particularly concerning precise timing. The phenomenon of prepulsing, discussed in detail in 11.3, originates from the detection of transmitted light by d_1 . The corresponding anode signal is undersized and arrives earlier than a signal originating from the photocathode; the time difference is $t_{k-d1} - t_\gamma$, where t_{k-d1} is the photo-electron transit time, and t_γ that for a photon. Although prepulses are more discernible in large-diameter PMTs, through their more favourable time separation from the main pulse, their occurrence is infrequent at a level of 0.1 to 1% of the photoelectron signal rate. Prepulses contribute to anode signal under normal PMT operating conditions but they make no contribution to a QE measurement under photodiode test conditions. This discussion highlights one of the many problems associated with making absolute light measurements: the calibration

method must mimic the application regarding polarization, wavelength, and incident angles.

It is clear from 2.3.4 that light sources can be calibrated by national laboratories with an uncertainty approaching a fraction of 1%. This also applies to certain detectors such as PIN diodes and PMTs operated in the photodiode mode, but considerable uncertainty arises in attempting to calibrate the anode of a PMT, although this is where the user's interest lies. Standards laboratories will not provide this form of calibration. Among other considerations, anode calibration requires knowledge of the PMT gain, which, as discussed in 5.10, carries uncertainties of the order of a few per cent. Collection efficiency, covered in Chapter 10, has similar uncertainty, while a multiplicity of optical effects, covered in this chapter, introduce perhaps greater uncertainty. The sought after absolute calibration of PMT output, in terms of amps per watt or picocoulombs per photon, remains elusive and restricted to an uncertainty of perhaps 5%.

REFERENCES

- Ankowski, A. *et al.* (2006). Characterisation of ETL 9357FLA photomultiplier tubes for cryogenic temperature applications. *Nucl. Instr. and Meth. in Phys. Res. A*, **556**, 146–57.
- Araújo, H. M. *et al.* (1997). Low temperature performance of photomultiplier tubes illuminated in pulse mode by visible and vacuum ultraviolet light. *Rev. Sci. Instr.*, **68**, 34–40.
- Araújo, H. M. *et al.* (2004). Low-temperature study of 35 photomultiplier tubes for the ZEPLIN III experiment. *Nucl. Instr. and Meth. in Phys. Res. A*, **521**, 407–15.
- Bell, K. W. *et al.* (2001). The development of vacuum phototriodes for the CMS electromagnetic calorimeter. *Nucl. Instr. and Meth. in Phys. Res. A*, **469**, 29–46.
- Born, M. and Wolf, E. (1999). *Principles of Optics*. Cambridge University Press, Cambridge.
- Canfield, L. R. (1998). 'Photodiode detectors'. In *Vacuum Ultraviolet Spectroscopy II*, ed. J. A. R. Samson and D. L. Ederer, Volume 32, Experimental Methods in the Physical Sciences, Elsevier, Amsterdam, 117–38.
- Charman, W. N. (1969). Properties of photomultipliers with caesium telluride photocathodes. *J. Sci. Instr. (J Phys E) Ser. 2*, **2**, 157–61.
- Fraser-Mitchell, J. and Wright, A. G. (1990). Contribution of photocathode nonuniformity to energy resolution in NaI(Tl) scintillation detectors. *Nucl. Instr. and Meth. in Phys. Res. A*, **288**, 429–38.
- Gunter, W. D., Grant, G. R., and Shaw, S. A. (1970). Optical devices to increase photocathode quantum efficiency. *Appl. Opt.*, **9**, 251–7.
- Hallensleben, S., Harmer, S. W., and Townsend, P. D. (1999). Limitations for enhancement of photomultiplier quantum efficiency through multiple total internal reflection. *J. Phys. D: Appl. Phys.*, **32**, 623–8.

76 Photocathodes

- Hallensleben, S., Harmer, S. W., and Townsend, P. D. (2000). Optical constants for the S20 photocathode, and their application to increasing photomultiplier quantum efficiency. *Opt. Commun.*, **180**, 89–102.
- Hoenig, S. A. and Cutler, A. (1966). Polarization sensitivity of the RCA 6903 photocathode tube. *Appl. Optic.*, **5**, 1091–2.
- Ichige, M. et al. (1993). Operating characteristics of photomultipliers operated at low temperature. *Nucl. Instr. and Meth. in Phys. Res. A*, **327**, 144–7.
- Janecek, M. and Moses, W. W. (2008a). Optical reflectance measurements for commonly used reflectors. *IEEE Trans. NS*, **55**, 2432–7.
- Janecek, M. and Moses, W. W. (2008b). Measuring light reflectances of BGO crystal surfaces. *IEEE Trans. NS*, **55**, 2443–8.
- Jones, D. P. (1976). Photomultiplier sensitivity variation with angle of incidence on the photocathode. *Appl. Opt.*, **15**, 910–14.
- Kobayashi, M. (1991). Transmittance and radiation resistivity of optical glues. *Nucl. Instr. and Meth. in Phys. Res. A*, **305**, 401–405.
- Köhler, R., Goebel, R., and Pello, R. (1995). Results of an international comparison of spectral responsivity of silicon photodetectors. *Metrologia*, **32**, 463–8.
- Larason, T. C., Bruce, S. S., and Cromer, C. L. (1996). The NIST high accuracy scale for absolute spectral response from 406 nm to 920 nm. *J. Res. Natl. Inst. Stand. Technol.*, **101**, 133–40.
- Lay, M. D. (1996). Parameterization of the angular response of the R1408 Hamamatsu eight-inch photomultiplier tube to be used in the Sudbury Neutrino Observatory. *Nucl. Instr. and Meth. in Phys. Res. A*, **383**, 485–94.
- Lay, M. D. (1997). Parameterization of the angular response of 2.54-cm photomultiplier tubes. *Appl. Opt.*, **36**, 232–39.
- Liepmann, M. J., Boehm, L., and Vagish, Z. (1992). Gamma radiation effects on some optical glasses. *SPIE*, **1761**, 284–95.
- McMillan, J. E. and Reid, R. J. O. (1989). The low temperature performance of fast photomultipliers. *J. Phys. E: Sci. Instr.*, **22**, 377–82.
- Meyer, H. O. (2010). Spontaneous electron emission from a cold surface. *EPL*, **89**, 58001.
- Moorhead, M.E. and Tanner, N. W. (1996). Optical properties of an EMI K₂CsSb bialkali photocathode. *Nucl. Instr. and Meth. in Phys. Res. A*, **378**, 162–70.
- Motta, D. and Schönert, S. (2005). Optical properties of bialkali photocathodes. *Nucl. Instr. and Meth. in Phys. Res. A*, **539**, 217–35.
- Murray, R. B. and Manning, J. J. (1960). Response of end window photomultiplier tubes as a function of temperature. *IRE Trans. NS*, **7**, 80–6.
- Nakamura, K. et al. (2010). Latest bialkali photocathode with ultra high sensitivity. *Nucl. Instr. and Meth. in Phys. Res. A*, **623**, 276–8.
- Ohno, Y. (1997). Improved photometric standards and calibration procedures at NIST. *J. Res. Natl. Inst. Stand. Technol.*, **102**, 323–31.
- Ohno, Y. (1999). Trilateral intercomparison of photometric units maintained at NIST (USA), NPL (UK), and PTB (Germany). *J. Res. Natl. Inst. Stand. Technol.*, **104**, 47–57.

- Piétri, G. (1994). *Photomultiplier Tubes: Principles and Applications*. Philips Export BV, Amsterdam.
- Persyk, D. E. et al. (1976). Photomultipliers for high-temperature applications. *IEEE Trans. NS*, 23, 185–8.
- Rambo, B. E. (1964). *Improved Long Wavelength Response of Photoemissive Surfaces*. Air Force Technical Report ALTDR 64-19. Air Force Avionics Lab, Wright-Patterson AFB, Ohio.
- Singh, A. S. and Wright, A. G. (1987). The determination of photomultiplier temperature coefficients for gain and spectral sensitiviy using the photon counting technique. *IEEE Trans. NS*, 34, 434–7.
- Sommer, A. H. (1980). *Photoemissive Materials*, Robert. E. Krieger Publishing, New York.
- Spicer, W. E. (1961). The production of pairs in semiconductors by low energy electrons. *J. Phys. Chem. Solids*, 22, 365–70.
- Timan, H. (1976). Optical characteristics and constants of high efficiency photo-emitters. *Rev. Tech. Thomson-CSF*, 8, 49–84.
- Zalewski, E. F. and Geist, J. (1980). Silicon photodiode absolute spectral response self-calibration. *Appl. Optic.*, 19, 1214–6.
- Zheng, X. et al. (2000). High-accuracy primary and transfer standards for radiometric calibration. *Chinese Sci. Bul.*, 45, 2009–13.

3

The optical interface to PMTs

3.1 Introduction

The performance of a PMT-based detector is usually limited by the number of photons incident on the photocathode and it is therefore essential that most available light be collected. This requirement would appear to be automatically satisfied for those applications in which the light is incident as a focussed or collimated parallel beam. Configurations for enhancing the effective QE are covered in 3.5 and 3.13. Optical systems may be broadly classified as imaging or non-imaging, but we deal with non-imaging applications only. Consequently, there is no requirement for image preservation and the only concern is that of optimizing light collection. Advantage is taken of the removal of this restriction in two aspects of optics that are particularly relevant to PMTs. The theory of light concentrators, eloquently developed by Welford and Winston (1978) is one, and the other concerns adiabatic light guides, first described by Garwin (1952). Underlying any attempt at concentrating or guiding light is subject to a fundamental limitation referred to variously as étendue, extent, acceptance, Lagrange invariant, or Liouville's theorem (the latter referring to conservation of phase space in classical mechanics).

The chapter begins with consideration of two parameters of fundamental importance to scintillators and other optical systems: light attenuation and reflection coefficients. Methods for collecting light from point sources follow, highlighting the general inefficiencies of lens systems. A significant portion of the annual production of PMTs is used in the manufacture of radiation detectors, based on low- and high- Z scintillators, in both sheet and cylindrical geometries. Steps that may be taken to achieve optimal performance throughout the energy range are outlined. Light concentrators, based on the methods of Welford and Winston, offer a means for gathering more light from a source through the use of a light concentrator. An integrating sphere is an example of the use of non-imaging optics based on diffuse reflection. Although a relatively minor application for PMTs, its mode of operation bears similarities to an important class of organic and inorganic scintillation detectors, and therefore warrants consideration. This analysis shows the importance of high reflectivity and stable material to efficient light gathering. The nature of the output from an ideal scintillator, in the form of a parallelepiped, is scale invariant. The concepts of escape cones and trapped light are covered in some detail because their application is fundamental in achieving efficient scintillation counter design. The same theory is applied in designing light guides for lossless transmission. This is accomplished by application of the adiabatic theory first enunciated by Garwin (1952). Ideal efficiencies can be calculated for various parallelepiped scintillator configurations, assuming no losses from bulk attenuation and reflection. These examples highlight the inefficiency in light collection for non-optimal arrangements: only $\sim 0.01\%$ of the available light is collected when viewing one of the large-area faces of a parallelepiped with a single PMT, for

example. Efficiency can be improved considerably by collecting light from one, or both, of the smallest area faces; the incorporation of adiabatic light guides provides the opportunity to raise this figure to 50%. The author has measured the light output of various large-area scintillation configurations, verifying that the order of magnitude predictions are substantially correct. There is a class of large-area detectors that utilizes the elegant properties of wavelength-shifting (WLS) light guides, proposed by Shurcliff (1950). The first practical realization is attributed to Barish *et al.* (1978) based on WLS bars; this was followed by Artikov *et al.* (2006), who used thin optical fibres. Equally elegant is the system for extracting light from small thin tiles (typical dimensions: $10 \times 10 \times 1 \text{ cm}^3$) incorporating a small diameter thin fibre laid in a groove cut into scintillating material.

3.2 Light attenuation and reflection coefficients

3.2.1 Light attenuation length, γ

These are key parameters bearing on the performance of a scintillation counter, and yet published information is limited. The deficiency of data for scintillators reflects the practical difficulty in making such measurements. Care is necessary in interpreting quoted attenuation lengths: the expression, bulk attenuation length, as measured by Artikov *et al.* (2006), for example, refers to the attenuation in the bulk material only, excluding reflections of any kind. Most measurements reported in the literature concern the effective transmittance, λ_{eff} , of scintillator materials and include contributions from reflected light; this parameter is sometimes referred to as the technical attenuation length (TAL). Transmittance (T) and attenuation length (γ) are related as follows: $T = I/I_0 = \exp(-\gamma x)$. The method described by Barton *et al.* (1964) for liquid scintillator gives an insight into the difficulties referred to above. Saint-Gobain (Bicron) adopts a technique similar to this one to deduce the attenuation length for plastic scintillator strips of various lengths (Saint-Gobain 2016). Jordan *et al.* (2003) give figures for a range of plastic scintillators, with and without reflective wrapping. It is clear from the experimental results that γ ranges from about 100 to 470 cm for organic scintillators; these figures refer to attenuation of the scintillator's own light, of course. Limited information is available for inorganic scintillators. Studies have been reported by Vittori *et al.* (2000) on crystal pillars of YAP(Ce) and CsI(Tl), ranging in length from 3 to 28 mm, and in cross section either 0.6×0.6 or $1 \times 1 \text{ mm}^2$. They find that the light collected by the crystal of longest length is a factor of 20 lower than that for the shortest. This is attributed by the authors to light trapping caused by the high refractive index of the crystal material, plus a contribution from reflection losses. Zhu *et al.* (1996) report on TAL measurements for PbWO_4 with γ ranging from 156 to 259 cm. Chen *et al.* (2005) quote 55 cm for LYSO, 84 cm for BGO,

and 130 cm for LSO. Vilardi *et al.* (2006) give results for $0.32 \times 0.32 \times 10 \text{ cm}^3$ YAP(Ce) and LYSO bars, for a variety of cladding materials: 20.8 cm for bare YAP(Ce), and 42.0 cm for LYSO. This informative paper also examines the effect of cladding on timing and position resolution.

3.2.2 Reflectance coefficients, ρ

The nature of reflection processes needs elaboration. Ideally, reflections are either diffuse, or specular. Although, certain diffuse reflectors have a significant specular component at glancing angles. Diffuse reflections are produced from rough surfaces, or from scattering within a material, an example of which is reflection from PTFE or TiO_2 . Light penetrates the surface, suffers scattering, and is ejected at random angles. This is a bulk effect, which explains why some diffuse reflectors are only efficient if sufficient scattering material is provided, typically a thickness of 0.3 to 1.0 mm is necessary. A Lambertian surface is a special case of diffuse reflection in which the surface appears equally bright from all viewing angles, or equivalently stated, the reflected light is proportional to the cosine of the angle between the normal and the angle of observation; this is illustrated in Fig. 3.1 and is known as Lambert's cosine rule.

Small volume scintillators are invariably clad with reflective material, but the quality of performance depends critically on the nature of these materials. Janecek and Moses (2008a, b) carried out a detailed study of materials suited to scintillation counters. They have measured reflection coefficients, ρ , as a function of incident angle, and the results in Table 3.1 refer to all incident angles between 14° and 78° . Janecek (2012) provides further results, using an integrating sphere, for the measurement of diffuse materials.

The article by A Springsteen (2009) is a readable account that provides not only an historical sketch, but also practical details on the various reflectors for use with integrating spheres; most of these materials can also be used in scintillation counters. Ideally, coating materials should be highly diffuse (fully Lambertian), highly reflective over the wavelength range of interest, highly stable, easily applied, easily removed, and immune to attack from pollutants.

Magnesium oxide is one of the earliest materials used; it has a reflection coefficient of 0.95 for wavelengths greater than $\sim 500 \text{ nm}$ but its reflectivity drops rapidly below 400 nm—not ideal for scintillation counter purposes. Another traditional material is barium sulphate, with a reflection coefficient of 0.98 for wavelengths longer than 300 nm. This material was commercially available as Eastman 6080 white reflecting coating, comprising BaSO_4 powder, a PVA binder, water, and alcohol. It is currently available as Avian-B white reflective coating.

PTFE powder in a base of silicone vacuum grease has a reflection coefficient in excess of 0.99 for wavelengths exceeding 300 nm and it is Lambertian for reflected angles below 78° . Fragility and contamination are disadvantages. Sintered solid PTFE has high reflectance in excess of 0.99 for wavelengths beyond 300 nm and is available from several manufacturers such as SpectralonTM, FluorilonTM,

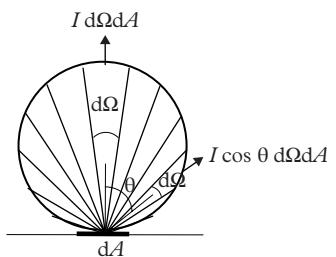


Fig. 3.1. Illustrating Lambertian emission. The number of photons emitted through each segment is proportional to its area.

Table 3.1 Reflection coefficients of commonly used materials (Janecek and Moses 2008b). All measurements refer to 440 nm. Enhanced Specular Reflector Material (ESR2) is an organic film made by 3M™ and marketed as VM2000, VM2002, and VM3000. The first set of materials listed in the table are diffuse reflectors, and the entries listed subsequently are specular ones.

Material	Reflection coefficient
Teflon® plumber's tape, three layers	0.992
Teflon® plumber's tape, four layers	1.000
PTFE	0.934
DuPont™ Tyvek® paper	0.979
Titanium dioxide paint	0.951
TORAY, Japan, Lumirror® E60LWRF, polyester	0.990
VM2000	0.944
VM2002	0.960
VM3000	0.890
Melinex®, DuPont™	0.969
Aluminium cooking foil	0.787

Op-Di-Ma™, and Zenith™. An interesting and surprising result reported by workers at NIST, Gaithersburg, MD, USA, highlights the susceptibility to contamination of PTFE from hydrocarbons, citing a case of damage caused by emission from engine exhaust (P-S Shaw *et al.* 2007). This same paper gives useful information on various aspects of integrating sphere usage and on fluorescence effects between 300 and 400 nm, the latter perhaps not a concern in scintillation counters.

Janecek and Moses (2008a, b) show that all diffuse reflectors, examined in their investigations, are not truly Lambertian because of the appearance of a specular peak at incident angles greater than 50°. At 80°, for example, about 20% of the light is specularly reflected and the amplitude of the true Lambertian component decreases accordingly. In the past, this aspect of specular reflection was not included in Monte Carlo simulations—arguably because its existence was unknown. In addition, the fact that bialkali photocathodes reflect a quarter of incident light, confirmed by Moorhead and Tanner (1996a) and others, tends to be ignored. Photocathode reflection coefficients are essentially flat at 20% for perpendicularly polarized light up to incident angles of 80°, and similarly for parallel polarization, together with a sharp peak at the onset of TIR. Monte Carlo calculations for scintillation counter design must include these reflection modes if a true simulation is sought.

Table 3.2 Reflection coefficients determined by J E McMillan, private communication, 2006. The first set of entries refers to diffuse reflectors and the second one covers specular ones.

Material	Reflection coefficient
PTFE plumbers tape, 1 layer	0.85
PTFE plumbers tape, 3 layers	0.95
PTFE plumbers tape, 10 layers	1.00
Engineering grade PTFE block	1.00
DuPont™ Tyvek® non-woven polypropylene	0.90
Expanded polystyrene	0.95
Millipore™ filter paper	0.87
Photocopier paper	0.86
Radiant light film 3M™ ESR2	1.00
Aluminised Mylar® sheet (Paperchase)	0.82
Aluminium cooking foil (dull side)	0.90
Aluminium cooking foil (shiny side)	0.82
PMT 9265Q	0.25

A word of caution in the use of commercial products for reflectors is appropriate: in particular regarding Tyvek®, Mylar®, and PTFE: these products were not developed for the purposes to which scientists apply them. Tyvek®, for example, was developed for heat insulation in domestic construction, and some of the grades are poor reflectors. Mylar®, a metallized polyester film, is available in two forms: one consisting of a thin aluminium layer on clear insulating material; this side is thus conducting and the other insulating with implications for grounding, either intentional, or unintentional. The existence of pinholes is serious where light tightness is required; manufacturers of beta-radiation probes, for example, overcome this problem by using two sheets to block ambient light leaks. Furthermore, the metallization is sufficiently thin to permit transmission of UV light from domestic fluorescent fittings. The other useful Mylar variant is metallized on both sides, thus overcoming the pinhole problem. Aluminium foil intended for cooking purposes has a matt and a shiny side and, counterintuitively, the matt surface seems to have the same or even higher reflectivity than the shiny side (see Table 3.2).

3.2.3 Optical coupling of scintillators

A typical detector comprises three elements: scintillator, light guide, and PMT, or, in many cases, only a scintillator and a PMT. How should the interfaces between

these elements be chosen to achieve the highest light collection? Three conditions must be satisfied: any medium sandwiched between a pair of interfaces must transmit over the full emission spectrum of the light, and without fluorescence; its refractive index is ideally intermediate between that of the two surfaces to be coupled; a minimum quantity of grease should be used to achieve a bubble free interface. The traditional method uses high-viscosity silicone grease; but silicone optical cushions, of thickness in the range 0.2 to 2mm are also effective and provide the additional benefit of a shock absorber. Thin cushions are preferred, to limit the escape of peripheral light (see 3.4, Fig. 3.5(b)). Clear epoxy optical cement is available from several manufacturers for making a permanent joint between elements.

Successful optical coupling with grease is a skill that is amenable to self-training. The problem is that the achieved quality, for example, when coupling an encapsulated NaI(Tl) crystal to a PMT, is out of view. The following dummy exercise is recommended for anyone new to the subject: that is, to couple a disc of plastic or glass to a PMT window using a rotary motion, noting how long it takes to achieve an interface free from bubbles and striations. Progress should be visually monitored to gain an indication of how best to achieve an acceptable result with the actual components. Large-diameter PMTs with plano-planar windows are particularly difficult to couple successfully: the window, although typically 7mm thick, bows under atmospheric pressure by a fraction of a millimetre at its centre. Not only does this take up a significant quantity of grease, but it also makes eliminating bubbles difficult to achieve. While it is feasible to obtain a flat window by grinding and polishing, manufacturers will need persuading to apply this to an evacuated PMT: possibly only to special order and at increased cost. Note that plano-concave windows, because of their structural strength, do not suffer from bowing. A further uncertainty with large-diameter PMTs is the release of coupling material over time—exacerbated at elevated temperatures. To ensure long-term stability, consideration should be given to operating with a small air gap, thus accepting a diminution in output signal and some loss in resolution (see 3.9, Fig. 3.19(e)).

3.3 Point light source

When a PMT and the source are contiguous, all light emitted through 2π solid angle is collected, and if a mirror is placed behind the source, as shown in Fig. 3.2(a), then in principle all light is collected. However, a portion of light incident on the window is reflected at the air–glass interface, and a higher proportion at the glass–photocathode interface. These so-called Fresnel losses are discussed in 2.4.2. Reflected light is not necessarily lost if the mirror dimensions and photocathode diameter are of sufficient extent. Few light emitters can be classified as purely point sources because there is invariably cladding in the source itself,

obscuring a portion of the emitted and reflected light. A small vial or an optical fibre inserted through a hole at the centre of the mirror defines a system that approaches the definition of a point source.

In many experimental arrangements, it is impossible to locate the source close to a PMT. An obvious but poor arrangement is to use a lens to focus the source on to the photocathode as in Fig. 3.2(b). It is poor because the solid angle, Ω , is invariably much less than 4π . Note that a 50mm diameter lens placed 100mm from a point source only collects 6.25% of the total light output. If such an arrangement is used, the lens should be a microscope objective with the source at its normal object position, including a concave mirror behind the source. Such lenses are available with f values as low as 2mm, and their use improves upon the poor geometry depicted in Fig. 3.2(b).

The alternative arrangement to the above, shown in Fig. 3.3, uses a solid light guide to couple a light source to a PMT. If the extent of the light source is smaller than the guide, which itself should be smaller in diameter than the PMT, then most of the light can be collected; this arrangement is improved by using optical contact at the light source (if appropriate) and at the photocathode window.

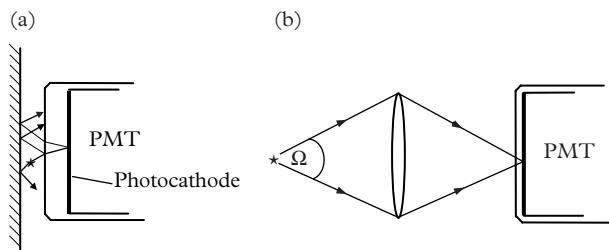


Fig. 3.2. (a) An acceptable arrangement for collecting light from a quasi-point source; (b) a low efficiency and poor method for collecting light over a solid angle Ω . The PMT window thickness in (a) has been exaggerated to aid clarity.

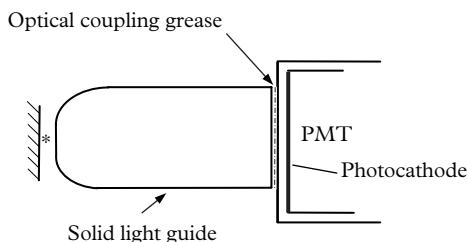


Fig. 3.3. The use of a solid light guide and a mirror to optimize collection. Light guides are discussed fully in 3.8.

3.4 Interfacing to small volume scintillators

Inorganic scintillators are characterized by high density and high refractive index n , while the reverse applies to the organics. The relationship between the refractive index and the critical angle θ_c is shown in Fig. 3.4: the result of a crystal–window interface of high n is a low critical angle, $\sin^{-1}(1/n)$. This extends collection of light randomly generated within a crystal, as a significant proportion is reflected from the interface. Clearly, γ and ρ are important where path lengths and the number of reflections are both extended. Perhaps reflecting the difficulty in making such measurements, there is a lack of attenuation data for both high- and low- Z materials. The parameter of interest is the effective attenuation length, TAL, discussed in 3.2.1; this ranges from 20 to 260 cm for high- Z materials, but it is critically dependent on intrinsic impurity levels. Plastic scintillator is particularly transparent to its own emission light, and TAL figures approaching 400 cm have been quoted. Further details on measured attenuation lengths and reflection coefficients are given in 3.2.1. Clearly, collection of light from small volume plastic scintillators is predictably more efficient, because of the favourable match at the scintillator–window interface.

A PMT window sometimes acts as an unintended light guide. Windows of 30 mm diameter, for example, may have any of the following configurations and thicknesses: plano-planar 1.25 mm thick; plano-concave of edge thickness 5.1 mm; thin meniscus 0.5 mm thick, or plano-prismatic. The author has made measurements on plastic scintillator of 25 mm diameter mounted in optical contact

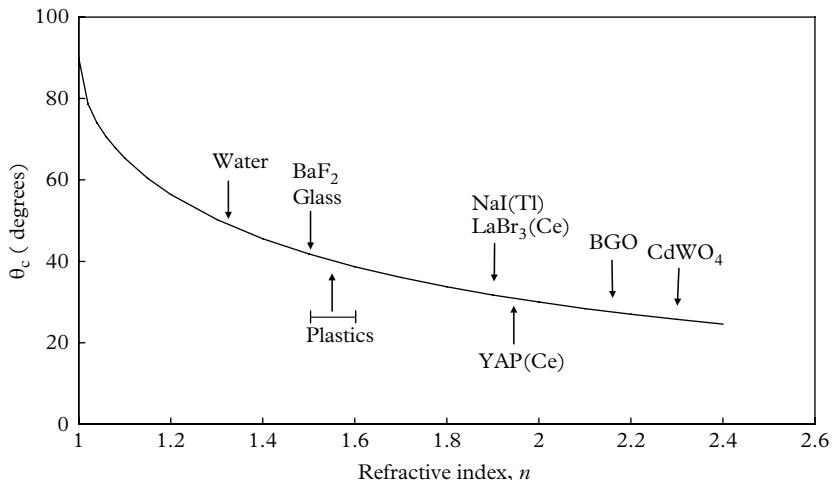


Fig. 3.4. Dense inorganic scintillators are notable for high refractive index, and hence low critical angle; BaF₂ is an exception.

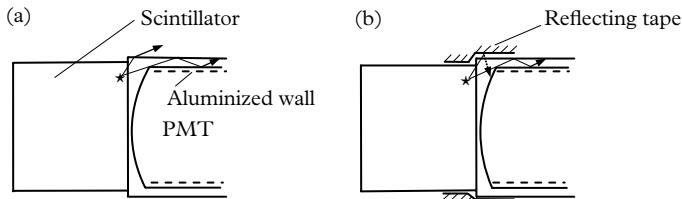


Fig. 3.5. A portion of light generated escapes through the edge of the window in (a). Light can also escape in this manner from scintillations originating within the bulk of the scintillator, particularly after several reflections off the walls. A portion of the escaping light can be directed to the photocathode by the addition of aluminium foil or PTFE tape, as shown in (b).

with a 30 mm diameter PMT with a plano-concave window. Wrapping a strip of aluminium foil around the cylindrical PMT–scintillator interface, in the manner illustrated in Fig. 3.5(b), increases the signal by 20%. PTFE in the form of plumber's tape may be more convenient, and it gives a similar result if sufficient tape is applied. An increase in signal of only ~2% results when the same procedure is followed with a PMT incorporating a 1.25 mm thick plano-planar window. The scintillator chosen for investigation is of smaller diameter than the window—a good rule to follow as the effective diameter of a PMT is always less than its physical diameter. Following this advice reduces peripheral light loss and leads to improved resolution. While statistical considerations relating to higher light collection contribute in the usual way, it is the removal of the position sensitive aspect of this light loss that brings benefit. The recommendation is to avoid using a PMT with a plano-concave window, unless speed of response is critically important. The existence of this particular escape path for light has been nicely demonstrated by D'Ambrosio *et al.* (1999), who replaced a quartz, plano-concave window ($n = 1.48$) of an HPD with one of intrinsic YAlO_3 (YAP). Signals derived from a BGO scintillator benefited from an enhancement of 1.78 times in the collected light, all of which is attributed to the better match of refractive indices of the two materials, 1.95 and 2.13 for YAP and BGO, respectively. High divergence experienced by rays at the crystal-quartz interface is reduced when substituting quartz with YAP, with two consequences: reduced light loss by the mechanism shown in Fig. 3.5(b) for escaping light and, a smaller fraction of light returned to the crystal.

Configurations shown in Fig. 3.6 illustrate the benefits derived from a solid light guide located between a cylindrical scintillator and a PMT. The surface of the scintillator is covered in diffuse reflector, except the exit face, and the light guide may be bare, covered in diffuse material, or wrapped in specular reflector. The first aspect of performance to discuss is the effect of using a PMT with non-uniform response across the window—a well-known defect in some PMTs. A point made in Chapter 2 is worth repeating here. Photocathodes usually have a uniform QE across the entire area of the window. Non-uniformity is evident at

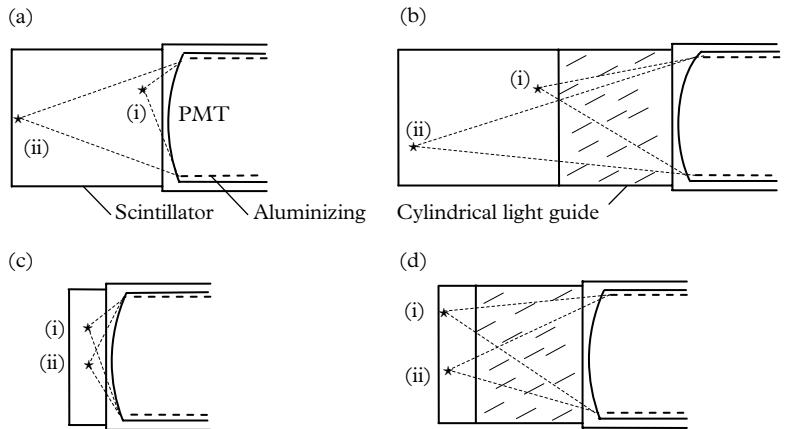


Fig. 3.6. The four illustrations show how the uniformity of response is improved by using light guides. Light reaching the photocathode directly from a cylindrical scintillator (i.e. without wall reflections) is contained within a cone included by the dotted paths. Optical coupling is assumed between all elements.

the anode primarily through the action of the multiplier. Whenever the term cathode uniformity is used, it is understood to refer to the variation in the amplitude of the signal seen at the anode, but referenced to the point of initiation on the photocathode. The arrangement in Fig. 3.6(a) is particularly sensitive to the location of the scintillation: a higher proportion of the total light generated strikes the photocathode directly in events of Type (i) compared with those of Type (ii). The local cathode sensitivity therefore has an effect for Type (i) but, for Type (ii), the direct light covers the entire photocathode surface reasonably uniformly. The inclusion of a light guide in Fig. 3.6(b) reduces the contribution from local hot spots thereby improving energy resolution. The same arguments apply to thin scintillators, Fig. 3.6(c) and (d), but even more so. These conclusions are supported by Monte Carlo and experimental studies by Fraser-Mitchell and Wright (1990) on 50 mm NaI(Tl) crystals ranging in height from 0.5 to 15.0 cm. Resolution with thin crystals is sensitive to cathode non-uniformity: the most uniform photocathode studied gave a resolution of 6.5% with NaI(Tl) compared with 7.1% for the worst sample. The studies also showed a tendency for light trapping in the far rim of a cylindrical scintillator, especially under low-reflectivity conditions ($\rho < 95\%$). Note that all the light, except the small amount absorbed in the bulk material of a scintillator, ultimately emerges in the ideal situation where ρ is unity.

3.4.1 Interfacing to scintillators of high Z

Inorganic scintillators such as NaI(Tl) are preferred for the detection of gamma rays, primarily because of their stopping power and high light yield. The discussion centred on Figure 3.6 highlights a positional dependence between the point of

initiation of a scintillation and the magnitude of the corresponding anode signal. There is also a dependence between the location of a radiation source relative to the position of the scintillator, but for present purposes we consider a point source of gammas positioned on the axis and in contact with the base of the crystal. The linear attenuation coefficient μ_o is a measure of the number of primary gammas that have reacted with the crystal in accordance with the $I/I_o \exp(-\mu_o x)$ formulation. The distance x after which half the incident gammas have interacted, but have not necessarily been totally absorbed, is $x = \ln(-0.5)/(\rho\mu_m)$, where ρ is the density of the material (1.03 for a plastic scintillator, and 3.67 g/cm³ for NaI(Tl)). The attenuation coefficients are related by $\mu_o = \rho\mu_m$, where μ_m is the mass attenuation coefficient and ρ the density. Data are given for μ_m by Evans (1955) for NaI(Tl), but it can be found in more recent publications, in addition to the internet. We deduce from the I/I_o relationship that $x = 0.032, 0.19, 2.47$, and 3.8 cm for ²⁴¹Am, ⁵⁷Co, ¹³⁷Cs, and ⁶⁰Co, respectively. These boundaries are illustrated in Fig. 3.7. For a point source and a 50 by 50 mm NaI(Tl) crystal, half the ²⁴¹Am gammas will have reacted within 0.32 mm of the base of the crystal, providing effectively a point source for all ²⁴¹Am captures. The location of a typical event is represented by the diagram marked (ii) in Fig. 3.6(a); in contrast, the interactions of the two higher-energy isotopes will be fairly uniformly distributed throughout the volume of the crystal, and inclusion of a light guide is beneficial in the latter case. An encapsulated sodium iodide crystal always includes a window: the 2R2, manufactured by Saint-Gobain (Bicron), has a window thickness of 6.4 mm for small crystals and up to 10 mm for larger ones (Saint-Gobain 2016). Ultra-low-background assemblies may include an unactivated NaI light guide, primarily acting as a gamma shield, but also providing optical benefits. The concern with the use of light guides is loss of light in a manner similar to the edge effect in PMT windows Fig. 3.5(b). This is not as crucial in massive detectors of high-energy gammas (energy > 10 MeV); loss of light is less crucial because photon statistics are not the dominant contributor to resolution at such high gamma energies.

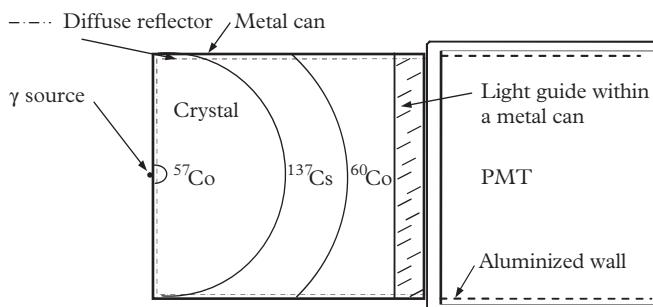


Fig. 3.7. An encapsulated 50 × 50 mm cylindrical NaI(Tl) crystal. Arcs refer to the depth at which half the gammas, from the source shown, have interacted. The window of the crystal acts as a short light guide.

3.4.2 Interfacing to scintillators of low Z

We are concerned in this section with plastic and liquid scintillators, primarily used in the fields of charge particle detection and in nuclear radiation probes. These scintillators have a density close to unity and consequently much lower stopping power than NaI(Tl). The corresponding figures, for the energies considered previously, for the half-distance x are 3.4, 4.1, 7.8, and 11.2 cm; consequently, points of origin for these scintillations are more uniformly distributed than those for high-density scintillators. There is thus minimal positional dependence in the initiation of events, compared with NaI(Tl), but a light guide of the form shown in Fig. 3.6(b) is still beneficial. The energy absorption of electrons is different from that of gammas because of their much shorter range. For 1 MeV electrons, the range is 4.4 mm in plastic, 2.1 mm in aluminium, and 1 mm in NaI(Tl), and all interactions are of the type depicted in Fig. 3.6(d). A thin crystal or plastic scintillator is therefore sufficient to stop all medium energy electrons. Uniform collection, over the entire photocathode area, is feasible by incorporating a light guide.

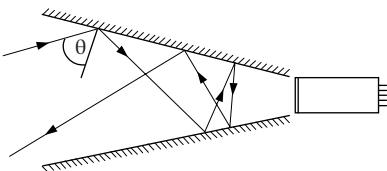


Fig. 3.8. A poorly designed wedge concentrator for wide-angle viewing, showing how a ray approaching at an angle of incidence, θ , is reversed without reaching the photocathode.

3.5 Light concentrators

3.5.1 Light concentrators: Theory

A concentrator is an optical means for enhancing light collection by a photocathode. Its efficacy is simply the ratio of the signal with optical enhancement to the signal with the enhancer removed. Non-imaging concentrators are sometimes described as light funnels, which unfortunately suggests an erroneous analogy with fluid flow. Light does not behave like fluid in a funnel—a common misconception. An example of this misunderstanding is the use of a wedge-shaped light concentrator in the false belief that light emanating isotropically from an area A can be concentrated onto a reduced area, a , for which a smaller-diameter and lower-cost PMT would suffice. Figure 3.8 illustrates that in a wedge-shaped arrangement some light can return without contacting the photocathode. In the section that follows, we examine the theory of light concentrators to show that the product of the area and $\sin^2\theta$ is conserved throughout the wedge. That is, as the area reduces in moving through the wedge, the angles of the rays expand to satisfy the stated invariant.

The function of the concentrator illustrated in Fig. 3.9 (in two dimensions only) is to accept a bundle of N rays of angular acceptance 2θ , incident on an entrance aperture $2a$; N' of these rays exit the optical system through a smaller aperture, $2a'$. The result is a concentration of optical power at the output but with a concomitant greater divergence in the exit ray bundle. The concentrator could be, for example, an integrating sphere or a convex lens, with the latter illustrated in

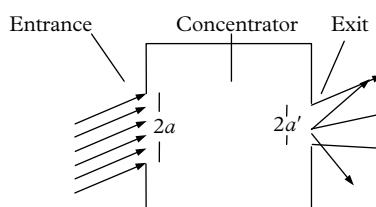


Fig. 3.9. Illustrating the general principle of light concentration.

Welford, W. T. and Winston, R. (1978). *The optics of non-imaging concentrators. Light and solar energy*. Academic Press, New York. Reprinted from Academic Press © 1978.

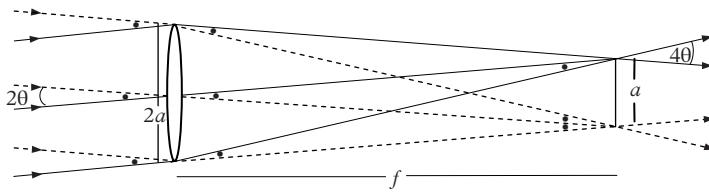


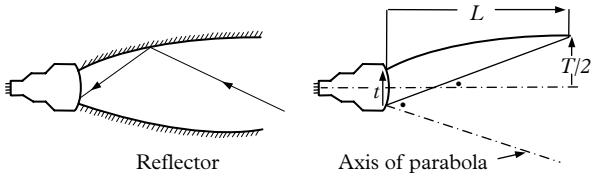
Fig. 3.10. Ray tracing for a convex lens viewing an object at infinity illustrates the basis of étendue: $a \times \theta$ is a constant throughout an optical system. All angles marked with a dot are taken as equal.

Welford, W. T. and Winston, R. (1978). *The optics of non-imaging concentrators. Light and solar energy*. Academic Press, New York. Reprinted from Academic Press © 1978.

Fig. 3.10. The usual small angle assumption in geometrical optics, which we now embrace, is that 2θ is small and hence $\sin 2\theta \approx \tan 2\theta \approx 2\theta$. We consider rays from a distant object with an angular acceptance of 2θ , constrained by an entrance aperture of diameter $2a$. For illustrative purposes, the image was chosen with a height of half the lens diameter. From the geometry of Fig. 3.10 we see that the product (input aperture diameter) \times (acceptance angle) is $2a \times 2\theta$ while, for the output, (image diameter) \times (exit angle) is $a \times 4\theta$, and numerically the same in both cases. This observation can be generalized to show that the product $a \times 2\theta$ remains constant throughout an optical system. Note, in three dimensions, the invariant is simply $(a \times 2\theta)^2$. The constant quantity is variously referred to as extent, étendue, acceptance, or a Lagrange invariant and it is customary to omit the factor of 4 in favour of just $a \times \theta$. Those schooled in classical mechanics will readily cite Liouville's theorem concerning conservation of phase space volume—hence the reference to Lagrange. Welford and Winston show that the prime parameter of any concentrator, involving two media of different refractive indices, is the light concentration ratio $(a/d')^2$, which has a maximum value of $n'/(n \sin 2\theta)^2$, where n and n' are the respective refractive indices, where applicable. From an experimental perspective, the concentration figure is simply the ratio of the light output with and without concentrator.

3.5.2 Light concentrators: Winston cone

Light is collected in Fig. 3.11 by a PMT with a window diameter of 12.7 cm (effective diameter 11 cm). The concentrator is a parabola rotated about the indicated axis, with an acceptance angle of θ_{\max} , generating a linear enhancement of $1/(\sin 16^\circ) = 3.6$. The authors do not quote the measured enhancement, although it will always be less than $(3.6)^2$ because of reflection losses in both the reflector and the air–glass–photocathode interface of the PMT. The procedure for arriving at the equation of the concentrator is given by the authors who note that the length of the funnel may be reduced by one-third without significant loss in efficiency. They also compare the performance of a truncated right circular cone,



3

Fig. 3.11. The earliest application of concentrators by Hinterberger and Winston (1966) using a hollow compound parabolic reflector to enhance the collection of Cerenkov light. The maximum acceptance angle is θ_{\max} , indicated by dots.

Welford, W. T. and Winston, R. (1978). *The optics of non-imaging concentrators. Light and solar energy*. Academic Press, New York. Reprinted from Academic Press © 1978.

which they say provides enhancement of $(1.7)^2$; according to Welford and Winston (1989) the enhancement for a paraboloidal mirror is lower. In a second paper, Hinterberger and Winston (1968) describe the combination of a solid concentrator in series with a hollow one, achieving an enhancement factor of $n^2/(\sin^2 \theta_{\max})$, where n is the refractive index of the enhancer, usually 1.5. Ouellette *et al.* (1992) discuss the use of concentrators in the form of thin films for application to hemispherical PMTs. The publication by Moorhead and Tanner (1996b) is particularly noteworthy for its detailed description of the design of an enhancer for 200mm hemispherical PMTs, immersed in water. Studying this paper and others by the same authors is highly recommended for anyone contemplating the design of a concentrator. Another remarkable design is that used in the MAGIC air Cerenkov telescope (ACT), employing a 25mm diameter hemispherical PMT and a hollow Winston cone of hexagonal geometry (Ostankov *et al.* 2000). The phenomenon of reflection from air–glass and glass–air interfaces is largely ignored in the design of enhancers because it is relatively small. There is a reduction in light transmission at all such optical interfaces amounting to a minimum of 4%, per interface, for light incident normally. The reflected component is more pronounced for a glass–photocathode interface where it is typically 20%. This reflected light is not necessarily lost but it should be included in all Monte Carlo simulations.

3.5.3 Integrating spheres

Integrating spheres have a history of over 100 years. The instrument consists of a hollow sphere, the inner surface of which is coated with diffuser of high reflectivity. There are at least two ports: an entrance port for the light and a viewing port for its detection. A hollow sphere randomizes the angular distribution of the input photon flux and can provide a measure of the total emission of lamps, LEDs, and lasers. A practical account of the subject may be found in the article by Lovell (1984) with a more theoretical technical guide currently available from Labsphere Inc. (2012). The ratio of the detected to the input flux (also known as the throughput) is given in (3.1) and is due to P-S Shaw *et al.* (2007):

$$\frac{\Phi_{\text{out}}}{\Phi_{\text{in}}} = \frac{\rho^2 A_2 [1 - (A_1 + A_2)]}{\rho [1 - (A_1 + A_2)]}, \quad (3.1)$$

where ρ is the reflection coefficient of the diffusing material, A_1 is the ratio of the entrance area to the total internal area of the sphere and A_2 is the ratio of the exit port to the total area of the inner sphere. If A_1 is sufficiently small and calling A_2 simply a , we have

$$\begin{aligned} \frac{\Phi_{\text{out}}}{\Phi_{\text{in}}} &= \frac{\rho^2 a (1 - a)}{1 - \rho (1 - a)} \\ &= \frac{\rho^2 a}{1 - \rho} \end{aligned} \quad (3.2)$$

$$\approx \frac{a}{1 - \rho}, \quad (3.3)$$

assuming ρ^2 is close to unity. These conditions are certainly satisfied when the area of the detector is considerably smaller than that of the sphere, and where a high-quality diffuser such as pressed PTFE is chosen.

3.6 Scintillators with diffuse reflecting walls

The expression in (3.3) was derived by Koechlin (1955) for scintillators enveloped in diffuse reflector. If a portion of the original light, a , escapes without reflection, then $(1 - a)$ hits the walls of which $(1 - a)\rho$ is reflected and $(1 - a)a\rho$ escapes, and so on. The total light escaping after n reflections is

$$\begin{aligned} S_n &= \sum_0^n a + a(1 - a)\rho + a(1 - a)^2\rho^2 + \dots + a(1 - a)^n\rho^n \\ &= \frac{a(1 - a^{n+1})}{1 - a} \end{aligned} \quad (3.4)$$

where $\alpha = (1 - a)\rho$. After a short time interval all the light will either have escaped or have been absorbed:

$$S_\infty = \frac{a}{1 - a} = \frac{a}{1 - \rho}, \quad (3.5)$$

the same result as in (3.3). The entity $1/(1 - \rho)$ in (3.5) effectively enhances the area of a PMT, a , and is known as the amplification factor. It can be considerable for a reflector of quality but for long-term stable operation it is important that the reflectivity does not deteriorate over time. The light collection process, S_n , is given in (3.4) and illustrated in Fig. 3.12 for different values of ρ . For ρ approaching

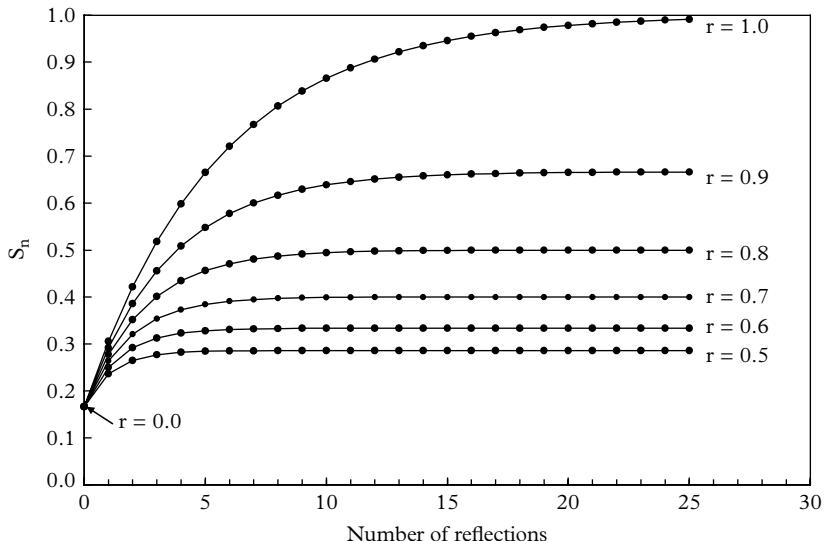


Fig. 3.12. Family of curves derived from equation (3.4) for a right cylinder with $\alpha = 1/6$. The point to note is that the major contribution to the collected light derives from a few reflections, where poor reflectors are used ($\rho < 0.7$). The opposite applies for ρ close to unity, which has obvious implications for timing.

unity, paths involving many reflections contribute to the light output although bulk attenuation in the scintillator will reduce the light collected. Also, long-term stability in output depends critically on maintaining ρ constant considering the sensitivity to the amplification factor. Applying diffuse reflector directly to a scintillator surface, as opposed to the housing containing the scintillator, provides a degree of interface protection from the environment, although light tight enclosures are generally hermetic.

A solid scintillator covered with diffuse reflector bears similarity to an integrating sphere but with two important differences: first, a significant proportion of light is absorbed by the scintillator unless its attenuation length is an order of magnitude greater than its dimensions and, second, scintillator-based detectors with spherical geometry are rare. However, based on the premise that a cube or cylinder is a reasonable approximation to a sphere, although a slab is not, we can use (3.5) to gain an insight into the process of capturing light from scintillators. The surface of a scintillator may be prepared in a number of ways: bare, covered with reflecting material, either specular or diffuse, or with the reflecting surface roughened. One of the flat faces is left untreated to allow the escape of light to a suitably disposed PMT or PMTs. The primary concern is the way in which photons are collected from a scintillator through multiple reflections. In the case of a right cylinder of NaI(Tl), the face coupled to the PMT constitutes 1/6th of the total surface area of the cylinder and consequently approximately 17% of the light

from a single scintillation strikes the PMT directly, and a further 17% of the 83% remaining reaches the photocathode after a single reflection, and so on. Light capture effectively terminates after about five bounces because the fractional exit area is relatively large (Fig. 3.12). A further simplification has been assumed in the last statement because scintillations originating close to the PMT window deliver approaching half the total light directly to the PMT, and $1/6^{\text{th}}$ only applies to events furthest from the window (see Fig. 3.6).

3.7 Light output from scintillator slabs

3.7.1 Escape cones

Large-area scintillators are those in which the surface extent is many orders of magnitude greater than the combined photocathode area of PMTs chosen for the detector. The use of a single PMT with slabs of up to 4 m^2 in surface area has been reported. To put this in context, a 51 mm PMT has a photocathode area of only $2 \times 10^{-3} \text{ m}^2$, which suggests that collecting sufficient light will be a formidable task, and yet detectors based on the principles of non-imaging optics have a long and successful history. Light generated by a single scintillation anywhere within a slab falls into two categories: that trapped by TIR, f_t , and the remainder, f_e , that escapes the scintillator. With a parallelepiped, once a photon is trapped it remains so; however, many TIRs take place until the photon is ultimately absorbed by the scintillator or by the photocathode. The demarcation between trapped and escaping light is the critical angle, θ_c , illustrated in Fig. 3.13 for a slab of refractive index 1.5.

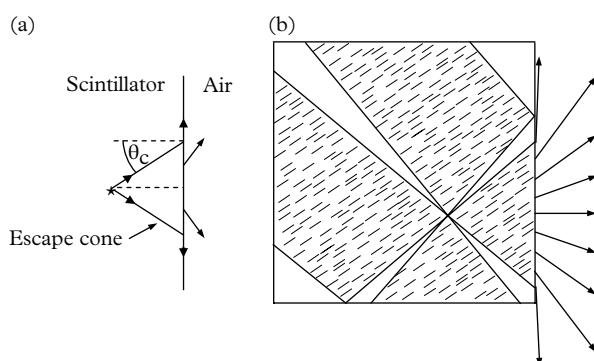


Fig. 3.13. (a) The definition of an escape cone. (b) Escape cones viewed from above showing four of the six cones for a slab of refractive index 1.5. All light emitted within the shaded cones escapes in the manner illustrated in (b) for the right-hand cone.

Light that escapes the face shown falls within a cone of solid angle, $\sin^{-1}\theta_c$. The fraction of the total light produced within a single escape cone is the same for each of the six faces and is given by the solid angle Ω , where

$$\Omega = \frac{1}{2}(1 - \cos\theta_c). \quad (3.6)$$

Also, the number of photons emitted in an escape cone is independent of the source location, on the assumption that the scintillator is totally transparent to its own radiation. The critical angle, θ_c , for glass or plastic, relative to air, is $\sin^{-1}(1/n)$, where n is in the region of 1.5, corresponding to $\theta_c = 41.8^\circ$, and thus, from (3.6), 12.7% of the light escapes through each of the six faces of a parallelepiped. The total proportion of light that escapes is $f_e = 6 \times \Omega$, or 76.4%, leaving a trapped fraction, f_t , of 23.6%. It follows from (3.6) that

$$f_t = 3\cos\theta_c - 2. \quad (3.7)$$

These equations may be expressed in terms of refractive index only using the identity

$$\cos\theta_c = (1 - \sin\theta_c)^{1/2} = \left(1 - \frac{1}{n^2}\right)^{1/2} :$$

$$f_t = \frac{3(n^2 - 1)^{1/2}}{n} - 2 \quad (3.8)$$

$$f_e = 1 - f_t = 3 \left[1 - \frac{(n^2 - 1)^{1/2}}{n}\right]. \quad (3.9)$$

It follows from Fig. 3.14 that no light is trapped in parallelepipeds with a critical angle greater than 48.2° , corresponding to $n < 1.34$. Four escape cones are shown in the plan view of Fig. 3.13. All photons travelling away from the point of scintillation and within the shaded area escape; this is illustrated for the right-hand escape cone only. Any ray travelling in the unshaded region of the diagram is totally internally reflected and cannot escape. The rays incident on the scintillator-air interface at an angle just short of the critical angle skim the face of the scintillator and escape. We ignore Fresnel reflections that at low angles of incidence amount to ~4% of the total, although there is a significant increase in this contribution as TIR is approached (see 2.4.2). This light is directed towards the opposite face contributing to the corresponding escape cone.

Some find the consequences of Fig. 3.13 counterintuitive: for example, the same number of photons is emitted from each face of a parallelepiped scintillator. The incorrect expectation is that more light should emanate from the two large-area faces. A simple demonstration involving a thin sheet of wavelength shifter readily dispels such a notion. When the sheet is excited with blue or ultraviolet light, the

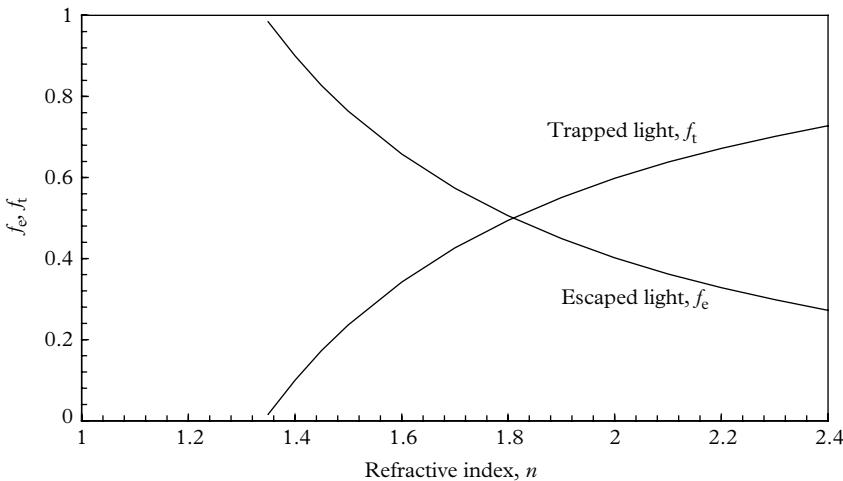


Fig. 3.14. Trapped and escaping light fractions as a function of refractive index. It follows from (3.8) that no light is trapped if $n < (9/5)^{1/2} = 1.34$, but note the high trapped fraction when n exceeds 2.

smallest area faces glow profusely, compared with the large-area ones, confirming that higher photon densities apply at the small-area faces. This observation points to the obvious but sometimes ignored generalization in large-area scintillation counter design: PMTs should be arranged to collect light from one or both of the two smallest area faces of a parallelepiped. Note that if a PMT of sufficient diameter to cover an entire face is optically coupled to the right-hand face in Fig. 3.13, all light incident on this face is collected; that is, one escape cone plus all the trapped light. The prime parameter, the light yield, can be estimated by using the simple but elegant theory developed in this section. The reader is also recommended the comprehensive theoretical investigation, with supporting experimentation, by Keil (1970). Barnaby and Barton (1960) were among the earliest to apply these formulations to large-area counters, some of which are now examined.

3.8 Light guides

Light guides serve two purposes: to transport light to a detector, and to improve uniformity of response over the entire detector volume. Distinguishing between a light guide and a light concentrator is essentially one of descriptive convenience. Propagating light by guides from a source to a photodetector is subject to the restrictions imposed by étendue, introduced in 3.5.1. There are two types of guide: solid and hollow. Solid guides are usually made of glass or Perspex, either circular

or rectangular in cross section. Included in this group are small-diameter fibres that offer the convenience of flexibility. Low-background apparatus may include an intrinsic form of scintillator as a light guide, while also serving to shield the scintillator from radiation emanating from the PMT window. Surfaces other than light exit areas may be bare, or covered with reflecting material. Hollow light guides are invariably specular.

3.8.1 Hollow guides with specular reflecting walls

Hollow, mirror-lined, light guides provide the means for improving uniformity in the same way as solid ones: that is, by distancing the PMT from scintillations that would otherwise occur close to the photocathode. The optimum length of the guide is a compromise between achieving uniformity and high light collection: noting that the average number of reflections per ray increases with the volume of the guide. A length of about 25% of the longest dimension would appear satisfactory, using material of the highest reflectivity, such as ESR2 (3MTM). There is no contribution from trapped light in this configuration, but light from a second escape cone contributes by placing a mirror on the opposite face to that carrying the PMT. Only a fraction of the initial flux traverses the light guide, for example, consider a slab of dimensions $72 \times 38 \times 3.8 \text{ cm}^3$ with the $38 \times 3.8 \text{ cm}^2$ face viewed through a hollow light guide of the same cross section. The calculation is straightforward for any given PMT and for present purposes a PMT of effective diameter 28mm is assumed. The collection efficiency, ignoring reflection losses, is in this case

$$\Omega = \frac{2 \times 0.127a}{A} = \frac{2 \times 0.127 \times \pi \times \left(\frac{2.8}{2}\right)^2}{3.8 \times 38} = 1.08\%.$$

Hollow guides with mirrored walls are relatively easy to construct, traditionally using aluminium (baking) foil or aluminized PET (Mylar®). Disadvantages of Mylar concerning pinholes and semi-transparency are mentioned in 3.2.2. In addition, a relatively low reflection coefficient applies. Higher-reflectivity materials, such as ESR2 (3MTM), although more costly, are ideal (see Table 3.1). Wrapping a scintillation counter with aluminium foil or Mylar® is common practice but it is important to follow the advice given by PMT manufacturers: anything in contact with the photocathode must be held at this potential. The length of a guide, constructed of low-reflectivity material, should be no longer than necessary. Hollow light guides are relatively inexpensive and are light in weight, and they do not produce Cerenkov light from relativistic charged particles.

3.8.2 Solid guides based on TIR

Solid (or liquid) guides generally rely on TIR, which in principle is 100% efficient for those rays striking the surface at incident angles greater than the critical one; this is 42° for $n = 1.5$. The favoured material is Perspex (also known as Lucite,

Plexiglas, or acrylic) but sometimes glass or quartz is chosen for low-background scientific experiments. All these materials function well in practice provided the surface is unspoilt. This proviso is important: there must not be any material touching the surface and it must be free from scratches, dust, and finger marks. Any attempted polishing should involve movement along the axis of the guide only. Particularly for long thin guides the quality of the surface is extremely critical and even the most careful polishing never seems to restore the surface finish to its original state. Cladded fibre-optic light pipes of small diameter are unaffected by the above considerations because the transmitted light is constrained within the bulk material, through the deployment of intrinsic graded index of refraction material (this is discussed further in 3.12). Although the bulk attenuation length in Perspex is several metres, the effective attenuation length (also referred to as the TAL), which allows for the different reflection modes, is less than the bulk.

Traditional light guides are sometimes wrapped in reflective foil in the mistaken belief that escaping light can be steered towards the photocathode. The angles of all such rays incident on the wall of the light guide are below the critical angle, therefore requiring more reflections to reach the photocathode than those propagating by TIR. Furthermore, light conveyed by TIR will be frustrated should the reflecting surface touch the light guide. There is another, even more futile, procedure of wrapping the light guide with crumpled foil; this is ineffective for the reasons already given. Aluminium foil is sometimes used to exclude external light, in which case at least two layers should be used to reduce the effect of pinholes.

3.8.3 Adiabatic light guides

It was shown in 3.5.2 that the photon intensity of collimated light could be enhanced by incorporating a compound parabolic reflector. The adiabatic principle enunciated by Garwin (1952) applies when the source of light is diffuse: ‘one can conduct essentially all the light travelling by TIR along a light pipe of any cross-sectional shape into another light pipe of any cross-sectional form, so long as the cross-sectional areas of the two light pipes are equal’. This applies to a light guide that changes shape continuously between its entrance and exit faces. A light guide is described as adiabatic because no light is lost in propagation, which requires that the change in shape must be gradual because severe localized changes allow light to escape. The term gradual is subjective and acceptable deformation limits can only be set by experimentation, Fukui *et al.* (1971), or by simulation. A typical transition is one from a rectangular to a circular cross section, as in coupling light emitted from one of the faces of a slab of plastic scintillator to a PMT. According to the manufacturer (Kuraray Co., Ltd), thin, graded-index light guides may be bent without significant loss of light, provided that the radius of curvature is ~ 100 times the guide diameter. Single refractive index guides, because of their higher critical angle, can tolerate more severe bending. Experimentation is the best means for establishing where the limits actually lie.

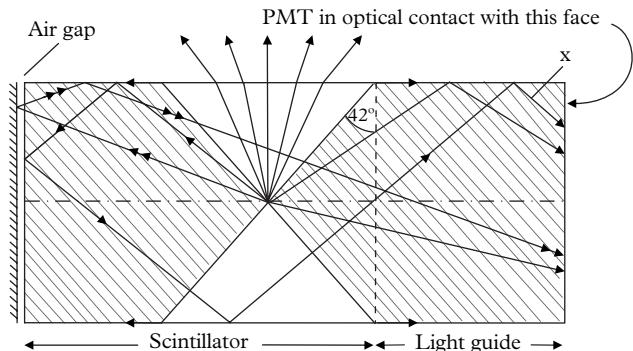


Fig. 3.15. A single light-emitting event in a parallelepiped-shaped scintillator coupled to a light guide of the same cross section and refractive index. All light launched into the shaded areas reaches the PMT, given there is no absorption in transit. For example, the ray marked 'x' is totally internally reflected four times before collection by the PMT.

The adiabatic principle enunciated above and its application is primarily for large-area thin scintillators. There are three distinct types of light guide: manifold, fishtail, and fibre. We first examine a parallelepiped scintillator and light guide combination with a manifold of just one element, covering the entire area of the smallest face. This is illustrated in Fig. 3.15, where the dashed line is the scintillator-light-guide interface. Optical coupling applies, making this interface invisible provided the refractive indices are the same. We assume that a PMT of sufficient diameter to cover the entire right-hand face is in optical contact with the light guide. The PMT collects all light incident on this face of the guide, regardless of whether it is escaping or totally internally reflected. A mirror, separated by a small air gap, is positioned on the scintillator face opposite to the location of the PMT.

An air gap is included to uphold TIR, while a mirror reflects escaping light. Light collected by a PMT comprises two escape cones in addition to all trapped light. The total light collected is $\Omega = 2 \times 12.7 + 23.6\% \sim 50\%$. If the effective diameter of the photocathode is d , and the length of the face concerned, L , the fraction of light collected, from a thin scintillator of thickness $\ll d$, scales as d/L . Approaching detection of half the light generated only applies to scintillators of negligible bulk light attenuation. Because the light guide is bare, transmission comprises direct and TIR components only. As mentioned previously, any attempt to collect the escaping light from the guide or scintillator by a foil covering tends to frustrate the light that would otherwise contribute. The largest area, plano-planar, PMT currently available has a diameter of 10 cm; by coupling such a PMT to a light guide with a width of less than 10 cm, still results in collection of 50% of the light generated. A wider guide would necessitate a multiplicity of PMTs to cover the entire exit face.

It is clear from the above discussion that the full photocathode area is underutilized by attempting to couple a rectangular cross section to a circular one.

Efficient use of the available photocathode area can be achieved by splitting a light guide into several parallel strips. The width of the central strip is usually chosen equal to the active diameter of the PMT with the remainder scaled down; strips comprising the manifold are twisted through 90° and bent, except for the central one (Fig. 3.16). A guide will approach adiabatic performance provided the change in shape, such as a twist or a bend, is gradual; this restriction affects the minimum length required of a guide. Investigations by Fukui *et al.* (1971), based on flexible light strips 80 cm long, report only 4% loss in transmission for 180° of twist over a length of 30 cm, and a loss of 3% for a 90° bend over a radius of 15 cm. This suggests that a manifold with 90° of twist and modest bending would be almost loss free. There are disadvantages with the manifold light guide: it is mechanically difficult to handle and mount. For long-term integrity, it is customary to couple the guide permanently to the scintillator and to the PMT with clear adhesive. A further possible disadvantage is the emission of Cerenkov light caused by the transmission of minimum ionizing particles through the guide.

Two types of fishtail guide are shown in Fig. 3.17: (a) transforms the rectangular cross section of a scintillator face to the circular geometry of a PMT window, and (b) transforms to a rectangular cross section approaching a square. The first is more costly to manufacture but makes full use of the photocathode area. Light is collected over the entire utilized face of the scintillator, and a mirror should be attached to the opposite face in the manner previously discussed. The operation is similar to the manifold type and the same rules concerning gradual deformation apply.

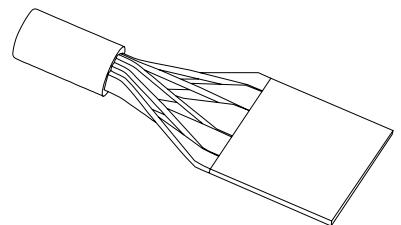


Fig. 3.16. A five-manifold adiabatic light guide, showing the interface to a PMT. It is important that the strips do not touch one another.

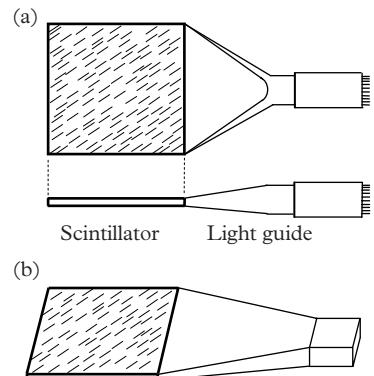


Fig. 3.17. Two types of fish-tail light guide: the one shown in (a) transforms a thin strip to a circular domain compatible with the PMT window dimensions, while the one shown in (b) transforms to a rectangle.

3.9 Light detection efficiency for scintillators

3.9.1 PMT viewing a large-area face

It follows from an earlier premise that collecting light in the way illustrated in Fig. 3.18 is inherently inefficient because of low photon density: the number per unit area scales with the inverse area of the face. Furthermore, since the PMT is remotely located, a quasi-inverse square law attenuation also applies. Without loss of generality, we assume a thin slab of scintillator (dimensions, 72 × 38 × 3.8 mm³) enclosed in a box of height 38 mm and lined with non-reflecting material. The reasons for this choice are the availability of such a slab for experimentation since it is a standard size for homeland security applications. Only a single escape cone contributes in this configuration, and therefore only a portion of the total light reaches a 30 mm PMT of photocathode area 6.2 cm². The light collected is, according to (3.6) and geometrical arguments, approximately

$$\Omega = \frac{0.127 \pi r^2}{A}, \quad (3.10)$$

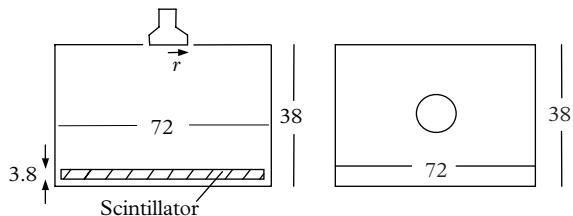


Fig. 3.18. A slab of scintillator of dimensions $72 \times 38 \times 3.8$ cm. The distance between the PMT and the scintillator should be at least 38 cm to achieve reasonable uniformity.

where A is the area of the upper part of the container (four sides and a lid), and r is the radius of the PMT. The proportion of light collected in this way is $0.127 \times 6.2 / 11096 = 7.1 \times 10^{-5}$. Barnaby and Barton (1960) warn that (3.10) provides an order of magnitude only, especially if applied to cylindrical and cuboid geometries. The collected light can be approximately doubled by placing a specular reflector beneath the scintillator, in optical contact, or otherwise. The additional contribution is reduced in proportion to the reflection coefficient of the mirror, and by attenuation in the scintillator. Because it is relatively thin, absorption losses in the scintillator should be small, as are reflection losses for a quality mirror.

Light collection remains very low and it is customary to attempt to collect from the four vertical faces of the scintillator. Swank (1954) notes that light specularly reflected within any of these four escape cones remains in this domain and can never reach the photocathode. Placing specular reflectors around the side faces is therefore useless and only the light emitted from the two horizontal faces has an opportunity for detection.

The use of diffuse reflector on the vertical sides allows some of this otherwise trapped light to escape, but its contribution is disappointingly small. The dominance of forward scatter, illustrated in Fig. 3.1, partly explains the failure to gather this light. Application of diffuse reflector to the lower face, in place of specular reflection, is effective: for now, preference for forward scatter is an advantage. Diffuser should also coat the top surface of the container, allowing a clear view for the PMT. The maximum improvement on the figure of 7.1×10^{-5} (a bare system) that can be achieved through the addition of reflector material is a factor of two to three, whichever reflector scheme is finally adopted (McMillan 2005). The theoretical amplification factor of $1/(1 - \rho) \approx 10$ for diffuse reflectors is clearly not achieved in practice because of both absorption losses in the scintillator and by light trapping. An improvement in uniformity of response also follows from the use of reflectors.

This type of counter has been successively used in air-shower experiments for the detection of minimum ionizing particles (MIPs). The number of photoelectrons generated by a MIP, such as an electron or muon, in traversing the scintillator may be estimated as follows:

the energy loss of an MIP is 1.5 MeV/cm of path

the energy liberated in a plastic scintillator of 3.8 cm thickness is 5.7×10^3 keV

plastic produces 8 blue photons/keV = 4.56×10^4 photons in total

the number of photons collected = $4.56 \times 10^4 \times 7.1 \times 10^{-5} = 3.24$ photons

the number of photoelectrons for a QE of 25% = 0.81 pe

A possible enhancement of three times with a reflector lifts the last number to 2.4 pe/MIP for a PMT of 30mm diameter. This is statistically insufficient for resolution purposes and points to the need for a PMT of at least 100mm in diameter (preferably hemispherical) which would generate a mean signal of the order of 27 pe, and hence a peaked response in the pulse height distribution.

McMillan (2005) describes an interesting variant of the large slab detector that overcomes the Swank limitation previously mentioned. He compared the performance of a $25 \times 25 \times 5$ cm³ slab with a mosaic of 25 units each of side 5 cm. In the first prototype, five faces of each small block were slightly roughened and covered in white paint, yielding an enhancement by a factor of 1.15, indicating only modest release of trapped light. Blocks were further disfigured in a second embodiment by drilling conical pits into five of the faces, followed by application of white paint. This yielded an improvement in light output by a factor of 1.93, which indicates that extreme disruption is required to release significantly more light. The reason for only a small improvement in the initial configuration is the dominance of forward projected light from the four vertical sides, consistent with the Lambertian angular distribution shown in Fig. 3.1. Track lengths in the small elements are considerably shorter than those in the original slab. This has the effect of reducing light losses by bulk attenuation; shorter track lengths also lead to a reduction in timing dispersion, which was McMillan's initial aim.

3.9.2 PMT viewing small-area face

Investigations concerning large-area configurations, outlined in Fig. 3.19, serve to verify (or not) the main assertions in this section. The results are based on a 30mm diameter PMT viewing the smallest area face of a $72 \times 38 \times 3.8$ cm³ slab of plastic scintillator. Measurements are based on the response to a ¹³⁷Cs source located centrally and on top of the large-area face and serving as a quasi-point source. The arrangement does not provide a true point source of light but rather one that is acceptable for present purposes. The shift in the pulse height spectrum, corresponding to the Compton edge, is taken as the relative change in light output. It is convenient to use the half height (on the high-energy side) of the spectrum as the measure, because, as is evident in Fig. 3.20, it is clearly delineated. The scintillator EJ200 has a quoted attenuation length for light of $\gamma = 250$ cm, which must be allowed for in assessing contributions via different paths between a scintillation event and its subsequent detection. Results

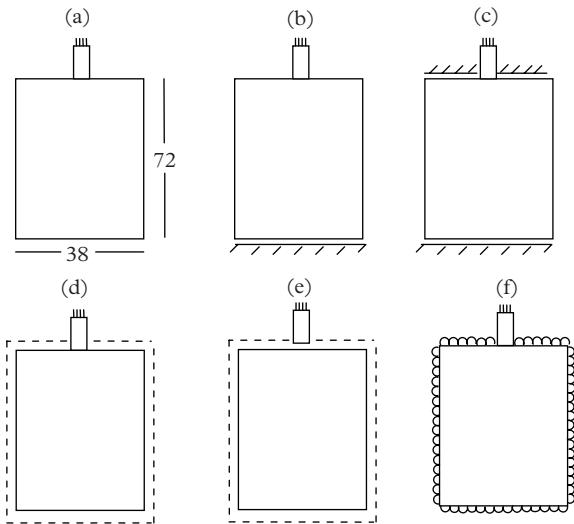


Fig. 3.19. Scintillator configurations examined by the author for relative light output from a ^{137}Cs source and a slab of plastic scintillator of thickness 3.8 cm.

are presented for various configurations including, domestic aluminium foil, glass-fronted mirrors, highly reflective organic films of ESR2, and diffuse TiO_2 paint.

A totally bare arrangement in Fig. 3.19(a) is taken as the reference for performance comparison; one escape cone, Ω_e , contributes in addition to the collection of an unpredictable fraction of trapped light, Ω_t . Although it is known that 23.6% of the light is trapped, we cannot calculate the fraction actually intercepted by the photocathode because this depends critically on the scintillator attenuation coefficient and path lengths involved. However, this contribution can be determined experimentally by noting the shift in the pulse height distribution when a small air gap is introduced between the PMT window and the scintillator, thereby removing contributions from TIR. It is found by this means that Ω_t accounts for about half the collected light in all the variants of Fig. 3.19. The predicted contribution from light in the single escape cone, $\Omega_e = 0.127 \times 6.2 / (3.8 \times 38) = 0.55 \times 10^{-2}$ and by including the trapped contribution, roughly 1% of the initial light should reach the photocathode. This is to be compared with collection of only $7.1 \times 10^{-3}\%$ when viewing the large-area face in Fig. 3.18.

The fraction of light detected in each of the following counter arrangements is expressed relative to the configuration shown in Fig. 3.19(a) and are indicated in parenthesis:

- (a) The scintillator is bare and light output is derived from one escape cone, Ω_e , plus a contribution from trapped light; the normalized light output is (1.00).

- (b) A mirror is included on the $38 \times 3.8 \text{ cm}^2$ face opposite to the PMT. The amount of light collected is increased by (1.25) and is insensitive to the choice between Mylar®, aluminium cooking foil, or domestic glass-fronted mirror. Note that, in the absence of attenuation in the scintillator, we would expect double the collection. However, light reflected off the mirror has to travel considerably further than the direct component. The quoted attenuation coefficient of 250 cm for the scintillator implies that one-quarter of the light is lost after a single traversal of the long dimension.
- (c) The addition of a second mirror increases the light collected to (1.57).
- (d) A bare scintillator, loosely wrapped in aluminium foil, has a relative efficiency of (2.00). Based on detecting a similar contribution from trapped light and the multiplier of 2.00 for aluminium foil, Ω follows as $0.53 \times 2 \times 2 = 2.1\%$. A figure of (3.44) is achieved with high-reflectance ESR2 film, corresponding to $\Omega = 3.6\%$. The relative light output from a counter with ESR2 film in this configuration compared to that using aluminium foil is (1.72)—a significant improvement.
- (e) The introduction of an air gap between the PMT and scintillator results in a loss of response from (2.00) to (1.00).
- (f) All faces are covered with the TiO_2 diffuse reflector, BC-620. This material has $\rho = 0.95$ and yet the relative light output is only (1.21); this is ascribed to the trapped light problem highlighted by Barnaby and Barton (1960) and McMillan (2005) and previously discussed.

The reflection coefficient for trapped light is theoretically 100%, and a portion of this light is detected by a PMT, after multiple reflections. However, the finite attenuation length, γ , of the scintillator means that some of this light is absorbed before reaching the PMT. Clearly, any estimate of the contribution from Ω_t is likely to be unreliable, and it is therefore best determined by Monte Carlo simulation or by experiment. Trapped light accounts for about half the collected light in all the configurations of Fig. 3.19. This is in accord with Monte Carlo simulations discussed in 3.9.4.

3.9.3 Photoelectron yield for scintillation counters

Gammas of energy 662 keV do not produce a photopeak in a plastic scintillator, because of the dominance of Compton scattering, but the backscatter peak at 477 keV is a convenient feature for determining the light output in terms of photons per keV or photoelectrons per keV. The figure of 8 blue photons/keV (or equivalently, 2 pe/keV) was assumed in 3.9.2 and applies here. Note the abscissa units in Fig. 3.20 are photoelectrons equivalent, achieved through calibration of an MCA with a single-photon source. The Compton peak for Configuration (d), with aluminium-foil wrap, occurs at 16 pe, with a corresponding efficiency of $\Omega = 16/(477 \times 2) = 1.7\%$, assuming 2 pe/keV. This is in agreement with the solid angle

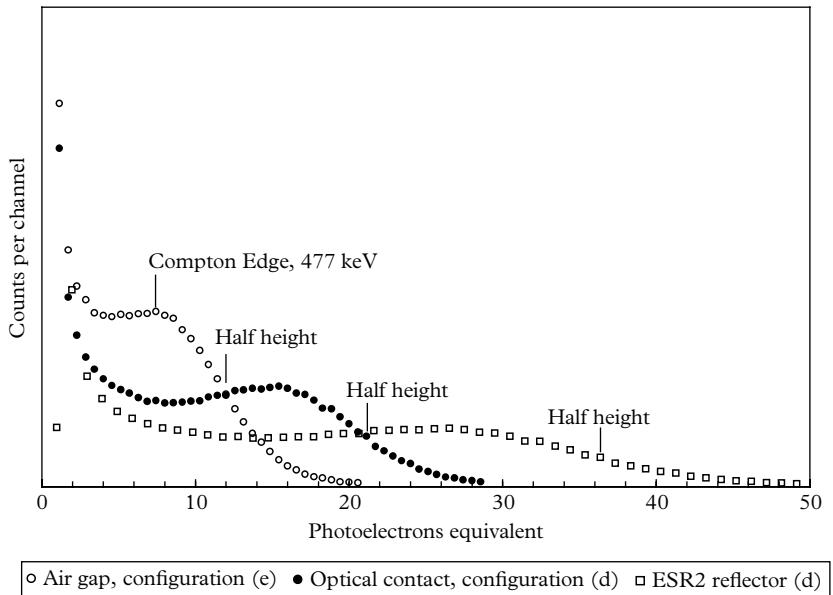


Fig. 3.20. ^{137}Cs spectra measured with a source located centrally and on top of a slab of plastic, configured according to Fig. 3.19(d) and (e). Referring to aluminium foil: •, 16 pe are detected for 477 keV energy deposited. Improved performance through the adoption of ESR2 organic film is immediately apparent, with a peak at 28 pe.

calculations and comparable to Ω_e for one escape cone, given previously as 0.53%. The spectra measured with a ^{137}Cs source placed at the centre of the slab are shown in Fig. 3.20 for three light collection modes. In the absence of optical coupling, only the light within two escape cones is collected. Optical contact, made with a 2mm cushion of clear silicone, doubles the light signal at the PMT.

Light collected from a slab of bare scintillator, with the dimensions assumed, and optically coupled to a 30mm PMT as in Fig. 3.19(a), is taken as the standard. The contribution made by the light in one escape cone, Ω_e , given previously, is $\Omega_e = 0.53\%$. The indication from Fig. 3.20 is that trapped light, Ω_t , makes a contribution similar in magnitude to Ω_e ; consequently, the total efficiency of an unclad counter should be of the order of 1%, and of about 2% when aluminium foil is attached. These efficiency figures are compatible with the figure of 1.7% deduced from measurements made with ^{137}Cs . It is always useful to perform order of magnitude calculations although losses through attenuation in the scintillator and reflection are essentially unknown. Although the discussion has centred on large-area detectors, the same considerations apply to other geometries and dimensions. A final point concerns the optimal thickness of a clear silicone pad, if used for optical coupling. Peripheral light gathering improves as the thickness is reduced, but if too thin, the efficacy as a shock-absorbing medium is compromised.

3.9.4 Monte Carlo simulations

Modelling based on Monte Carlo simulations, such as that performed by Geant and Detect98, is a powerful tool for predicting light collection efficiency, the uniformity of response, and timing. Parameters such as light attenuation, for the scintillator, reflection coefficient of wrapping material are readily included in the simulations. Investigations under the ALPS project carried out on large-area plastic scintillators at the Pacific Northwest National Laboratory are particularly noteworthy. Jordan *et al.* (2003) investigated the response of $127 \times 57 \times 5$ cm slabs of BC408 ($\gamma = 380$ cm) reporting, for example, predicted collection efficiencies of 3.8% for loose foil wrapping. This figure should be compared with 1.7% for the configuration in Fig. 3.19(d), noting some important differences in the parameters. A 50 mm diameter PMT was chosen for the ALPS study, and a 30 mm for the studies by the author on a smaller area (and thickness) scintillator of $72 \times 38 \times 3.8$ cm; the attenuation lengths of the two scintillators are significantly different at 380 cm and 240 cm, respectively. Nevertheless, quoted efficiencies are reconcilable within a factor of 2.

3.10 Uniformity of response in large-area detectors

An ideal detector is one in which the total collected light is independent of the point of initiation in the scintillator. The significance or otherwise of this aspect of performance depends on whether energy resolution is important, or, if the requirement is merely a signal for counting, for coincidence, or for anti-coincidence purposes. A straightforward way to map the scintillator response is to position a radioactive source at a set of locations, on a grid of ~ 10 cm spacing, referenced to the face of largest area. The quantity of light collected is known to be no more than a few per cent per event and hence a reasonably energetic gamma source should be chosen: ^{137}Cs is readily available in most establishments but any gamma emitter in excess of 0.5 MeV is satisfactory. A spectrum is recorded at each position and some feature, say the half height on the trailing edge of the spectrum, as before, may be taken as a measure of the relative light output (illustrated in Fig. 3.20). Note it is only necessary to make measurements over half the slab area, taking advantage of symmetry about the principal axis of the scintillator.

The response plot of Fig. 3.21 refers to the aluminium-foil-wrapped configuration shown in Fig. 3.19(d). The uniformity is remarkably good, considering the use of a single PMT, although enhancement is to be expected for events initiated close to the PMT. This is a well-known effect, which can be mitigated by various means. For example, using two PMTs sited on opposite faces leads to an improvement through superimposing two mirror image plots, such as those in Fig. 3.21. Measurements along the major axis of the scintillator are shown in Fig. 3.22 for

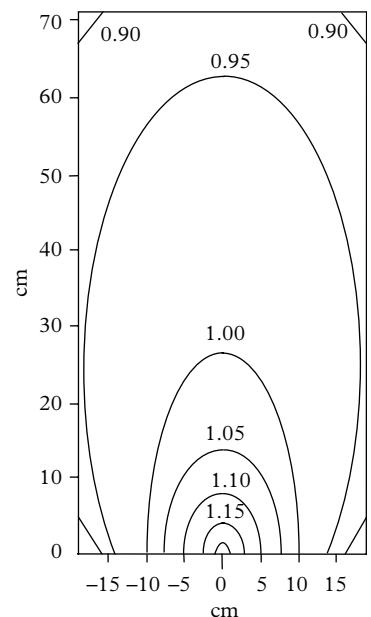


Fig. 3.21. Contour lines for light output referenced to the centre of the slab with the PMT located on the bottom face for configuration (d) in Fig. 3.19.

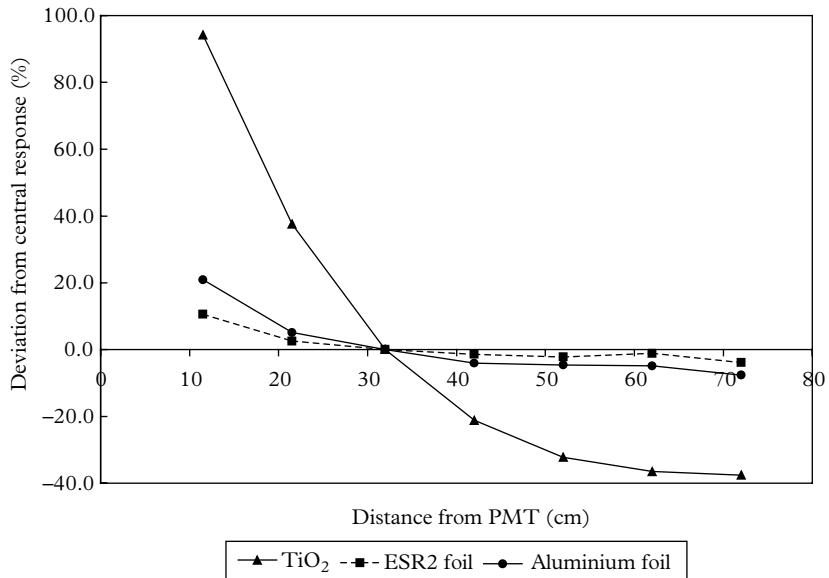


Fig. 3.22. Uniformity of response along the major axis of a large-area scintillator, configuration (d).

Configuration (d). Counters with aluminium foil or ESR2 reflector provide essentially the same performance with regard to source location; the curves in Fig. 3.22 suggest a means for improving uniformity using a solid light guide. The enhanced response for events close to the PMT, evident in Fig. 3.22, can be reduced by the attachment of a small slab of unactivated PVT to the scintillator. This should have the same cross section as the scintillator, with a length of the order of 20 cm in this case. The expected uniformity of response will be within $\pm 5\%$ over the entire active area of the scintillator. Scintillator manufacturers offer clear adhesives, of approximately the same refractive index as plastic, for attaching light guides to scintillators. Alternatively, a similar improvement is possible with an oversized length of scintillator, provided that the radiation can be confined to those areas of the scintillator sufficiently removed from the PMT—about 20 cm in this instance.

A diffuse reflector, as applied in Configuration (f), produces both poor collection and poor uniformity of response, highlighting its unsuitability for large-area detectors. The explanation for poor performance, peculiar to this configuration, lies in the random walk process followed by photons. With diffuse reflectors, a photon heading towards a PMT may at any reflection point reverse its direction, thereby reducing its detection probability. However, where specular reflections apply, a photon moving towards a PMT will continue broadly along its original direction after reflection, drawing closer to the photocathode, and hence with an increasing probability of detection. The severe non-uniformity for events initiated close to the photocathode should also be noted for diffuse reflection.

3.11 WLS light guides

Wave length shifting fibres (WLS) bars were first proposed by Shurcliff (1950) and subsequently used in very-large-area counters, such as those described by Barish *et al.* (1978; $150 \times 150 \times 4 \text{ cm}^3$ and $300 \times 300 \times 2.5 \text{ cm}^3$). The methods discussed so far for light collection would need either many PMTs to collect sufficient light from such large-area slabs, or cumbersome twisted adiabatic light guides of the type shown in Fig. 3.16. Neither of these proposals is satisfactory. The operation of a WLS counter is illustrated in Fig. 3.23(a). Light from one escape cone is intercepted by a WLS bar, of the same thickness and width as the scintillator, and doped with a fluor such as BBQ. Blue scintillator light is absorbed by the bar and readmitted isotropically at a longer wavelength in the region of 550 nm. This process is near instantaneous with a conversion efficiency of about 50% from blue to green light. A fraction of the green light escapes the bar but 16.7% propagates towards each end (3.11). The PMT is optically coupled to the bar, the other end of which carries a mirror to direct the light, which would otherwise escape, to the PMT. Not all the blue light is trapped by the bar on the first pass, but this component can be returned by loosely wrapping three of the long edges of the bar with foil. A bar can also be attached to a second edge of the scintillator and brought out to the same PMT, shown in Fig. 3.23(b). Mirrors on the remaining two long edges of the scintillator ensure that light from four escape cones is collected. For large-area configurations with dimensions comparable with the attenuation length in the scintillator, Barish *et al.* found it beneficial to use four shifting bars and four PMTs mounted at the corners. The level of doping in the BBQ bar is critical to attaining a long attenuation length for green light; this was 3 m for the material used by Barish *et al.* Bialkali photocathodes

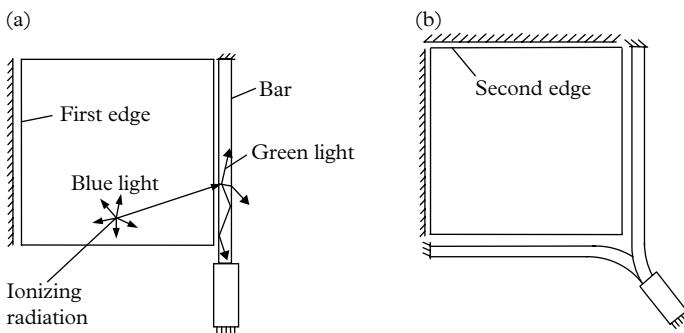


Fig. 3.23. Light collection based on WLS bars, which absorb blue light emitted from a slab of scintillator, followed by the emission of green light within the bar. The presence of a small air gap between the bar and the edge of the scintillator is essential to the optimal functioning of this detector.

provide the best match to blue scintillator light, but the spectral sensitivity of RbCs photocathodes is higher for green wavelengths. Viehmann and Frost (1979) describe a novel approach in which heavily doped wavelength shifter is applied as a 5 to 10 μm film to one face of a clear acrylic bar. In excess of 90% of incident light is absorbed while the shifted light traverses the film only on internal reflection; otherwise, propagation is in the clear part of the light guide. Attenuation lengths in coated bars are consequently longer than those attainable in uniformly doped ones.

3.12 Optical fibres

There are three types of commercial optical fibre: light transmitting, scintillating, and WLS. The light-transmitting types are light pipes for photons entering one of the end faces for transportation to the opposite end; they are available in circular and square cross section, with diameters ranging from 0.5 to 5.0 mm. Square cross-section, undoped fibres are simply long parallelepipeds, for which the analysis in Table 3.3 applies. Composite fibres (Fig. 3.24(b)) are those clad with a thin coating of clear material; the thickness is a few per cent of the core diameter, and the refractive index is chosen to be lower than that of the core (Fig. 3.25). The multi-clad form in Fig. 3.24(c) has inner and outer cladding with refractive indices of typically 1.49 and 1.42, covering a core of refractive index 1.59. The advantage of double cladding is that TIR takes place at the boundary of the two cladding materials. As this junction is internal, transmission is unaffected by bundling, handling, or by application of an opaque covering to the external surface. This is sometimes referred to as an extramural absorber (EMA), indicating that it is primarily intended to eliminate crosstalk from escaping light.

We want to know the fraction of light, f_{tr} , transmitted by a square fibre to a PMT in optical contact with the remote end face. This follows from previous considerations for light originating within a parallelepiped, as in Fig. 3.15. Light propagates, primarily by TIR, towards a PMT optically coupled to the fibre. The solid angle for light received by a PMT is

$$\begin{aligned} f_{\text{tr}} &= \frac{1}{2} \left(1 - \cos \left(\frac{\pi}{2} - \theta_c \right) \right) \\ &= \frac{1}{2} (1 - \sin \theta_c) \\ &= \frac{1}{2} \left(1 - \frac{1}{n} \right), \end{aligned} \tag{3.11}$$

Table 3.3 Calculation of θ_c and f_{tr} for square-section fibres. Propagation paths are shown in Fig. 3.25.

Type	θ_c	f_{tr}
Bare	$\sin^{-1}(1.00/1.59) = 39.0^\circ$	$\frac{1}{2}(1 - 1.00/1.59) = 18.6\%$
Single clad	$\sin^{-1}(1.49/1.59) = 69.6^\circ$	$\frac{1}{2}(1 - 1.49/1.59) = 3.14\%$
Double clad	$\sin^{-1}(1.42/1.49) = 72.4^\circ$	$\frac{1}{2}(1 - 1.42/1.49) = 5.34\%$

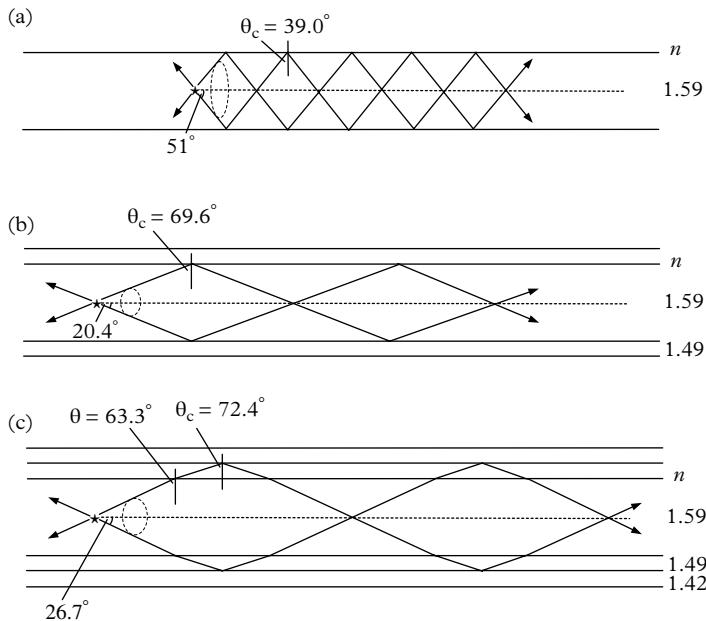


Fig. 3.24. Transmission of internally generated light down three fibre configurations. TIR is effective at the core-air boundary in (a), but in cladded fibres the important contributions derive from an internal boundary as in (b) and (c). These depictions apply to rectangular light guides and for light launched on axis for those of circular geometry.

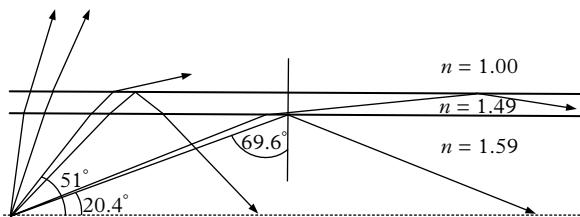


Fig. 3.25. The optics of a single clad fibre. TIR is effective at two optical boundaries but the internal one is the more important.

where $1/n = n_{\text{clad}}/n_{\text{core}}$; n_{clad} depends on the type of cladding material (unity in the absence of cladding). Note that f_{tr} depends critically on the ratio of the two refractive indices: the greater the difference in refractive index, the greater is the fraction transmitted. It is clear from the propagation modes depicted in Fig. 3.24 (b) and (c) that the light exiting a fibre is biased in the forward direction. Optical coupling at the PMT may be omitted with a small reflection loss at the air–glass interface. Circular fibres function similarly, with one important difference: f_{tr} depends on the point of origin of light emission, with up to twice the light propagating from points of initiation near the periphery of the fibre, compared with points near the axis.

Active fibres, doped with fluors, generate light within the fibre for immediate transportation by TIR. Scintillating fibres are fabricated with a polystyrene core that scintillates in either blue or green with a light-output efficiency similar to standard plastic scintillators. Blue and green scintillating fibres BCF-10 and BCF60 are available from Saint-Gobain (Bicron), and SCSF-38 and SCSF-3FH from Kuraray Co., Ltd. WLS fibres absorb blue scintillator light, of external origin, and shift it typically from 420 nm to 500 nm. The decay time of scintillators and shifters is of the order of nanoseconds and the attenuation length approaches three metres—both highly desirable properties. In some detector configurations, where light is transported several metres from the source, a combination of WLS and clear fibres (non-doped) are used to take advantage of the 10 m attenuation length of the latter. BCF92 and Y-11(200)M, from the aforementioned manufacturers, are well established in high-energy physics sampling calorimeters. Details of this type of detector are given by Wigmans (2000).

Unclad, round fibres have the highest efficiency, f_{tr} , of all variants in excess of 18.6% light transmission; single- and multi-clad round fibres have trapping efficiencies of 3.4% and 5.6%, respectively. Efficiency calculations for square-section fibres are given in Table 3.3, where it should be noted that square-section fibres have lower f_{tr} values than their round counterparts. Doped fibres are relatively transparent to their own emission wavelengths with transmission extending over several metres. Further information on this subject may be found in the technical literature of two companies: Saint-Gobain (Bicron) and Kuraray Co., Ltd.

3.12.1 Application of fibres to scintillating tiles

A calorimeter is a detector for measuring the energy loss of charged particles. In this instrument, some or all of the initial charged particle energy is dissipated in sheets of absorbing material, such as lead or iron. In the first generation of instruments, the absorber was interleaved with thin sheets of plastic scintillator, with each sheet viewed by one or more PMTs. In this manner, energy dissipation is sampled by the scintillators to provide a shower profile of the total energy loss. Some configurations employ the manifold type of light guide, while others use

wavelength shifter bars of the type described by Barish *et al.* (1978). Both of these techniques lead to bulky constructions, unsuitable for later generations of high-energy physics detectors. In addition, the ever-increasing energy attained by accelerators creates problems concerning radiation damage, which can be overcome by distancing all light detectors from the core of the calorimeter. A system of many, small, and thin tiles ($10 \times 10 \times 1 \text{ cm}^3$) of plastic scintillator with embedded WLS fibres has proved an efficient means of sampling and measuring energy loss. Many variants of the WLS technique have been reported in the scientific literature with a representative scheme described by Aota *et al.* (1999) in Fig. 3.26. They relied on an air-coupled WLS fibre placed in a groove machined in the plastic scintillator. The tile is traversed by MIPs in a track normal to the largest area face. Light moves outwards from these tracks towards the four lateral edges of the tile. A portion of the light is captured by a single WLS fibre, shifted to a longer wavelength, and piped to a PMT. If the scintillator is bare, the WLS fibre will collect about 10% of the light from the four escape cones, including a fraction of the trapped light; this totals no more than 20% and, when combined with a multi-clad fibre of efficiency 5.34% (Table 3.3), the overall efficiency reduces to about 1%. This is consistent with the yield quoted by Albrow *et al.* (1999) of 10 pe/MIP, noting that the PMT QE for green light is $\sim 10\%$. In plastic, dE/dx is 1.5 MeV/cm of track, producing ~ 8 blue photons/keV; this equates to a total of $\sim 1.3 \times 10^4$ photons. Assuming that all blue photons are shifted to a longer wavelength, then 10 pe (quoted by the authors) are derived from 100 green photons. The efficiency is therefore $100/1.3 \times 10^4$ or $\sim 1\%$ and is consistent with measurement. Asakawa *et al.* (1994) showed that the light yield is doubled with two turns of fibre, with little further gain from using more than three turns.

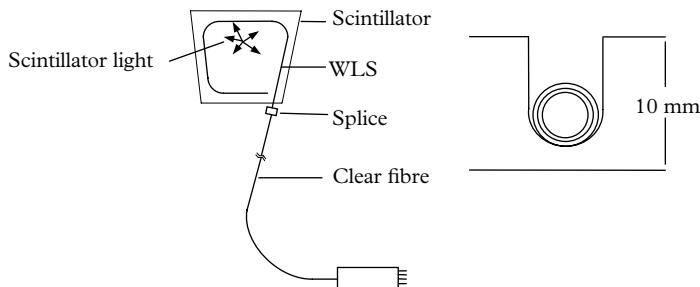


Fig. 3.26. A ‘sigma’ scintillation tile based on a composite WLS fibre and a clear fibre joined by a precision splice; the attenuation lengths in the fibres are 3 m and 10 m, respectively (Aota *et al.* 1999). The slot carrying a fibre is shown on the right, but some versions have a keyhole profile. A single multichannel PMT serves many tiles.

Aota, S. *et al.* (1999). Mass production of tile/fibre units for the CDF plug upgrade EM calorimeter. *Nucl. Instr. and Meth. in Phys. Res. A* 420, 48–61. Reprinted from Elsevier Science ©1999.

3.12.2 Use of fibres with large-area plastic scintillator slabs

Artikov *et al.* (2006) further developed the detector concept used by Barish *et al.* (1978), as detailed in Fig. 3.27, by replacing WLS bars with WLS multi-clad fibres of 1 mm diameter (Kuraray Y11, K27). These are fabricated in ribbon form of 20 elements and maintained in optical contact with the $230 \times 2 \text{ cm}^2$ face of a $230 \times 30 \times 2 \text{ cm}^3$ slab. One end of each fibre in a set is brought together and optically coupled to a small-diameter PMT; fibres at the other end of the ribbon are each terminated in a specular reflector. Contrary to the usual practice of collecting light from the smallest area face, these authors positioned the ribbon on the longest edge of the scintillator, providing both higher light yield and improved uniformity of response. An aluminium-foil strip on the face opposite the ribbon returns a further fraction of escaping light. One of the corners of the scintillator is removed to accommodate a small detector package, effectively within the footprint of the scintillator. This small perturbation of the geometry has no deleterious effect on performance, but provides an inherently more robust detector than those based on WLS bars. The mean pulse height for MIPs (cosmic ray muons, in this case) is quoted as 25 photoelectrons equivalent, or correspondingly 250 green photons; the initial number of blue photons generated for scintillator of 2 cm thickness is, assuming 8 photons/keV, 2.6×10^4 , and the efficiency is thus 1%.

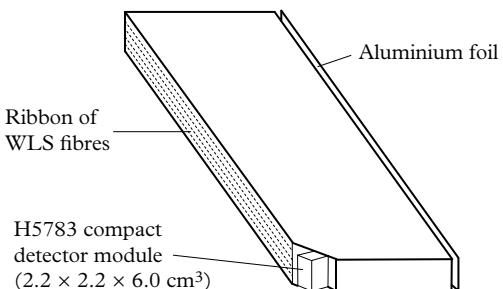


Fig. 3.27. A long slab of length 2.3 m of UPS923A-polystyrene-based, plastic scintillator with WLS fibre readout.

Artikov, A. *et al.* (2006). New generation large area muon scintillation counters with wavelength shifter fibre readout for CDF II. *Physics of Particles and Nuclear Letters*, 3, Issue 3, 188–200. Reprinted from Springer ©2006.

3.13 Techniques for enhancing light detection

3.13.1 Anti-reflection coatings

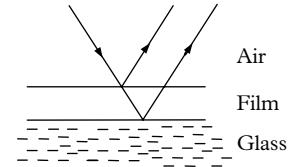
Reflections take place at a boundary between different optical media. According to the analysis in 2.4.2, 4% of normally incident light is reflected from the outer window surface of a PMT operated in air. This is undesirable in certain optical instruments, the familiar one being camera lenses, but there are means to frustrate the effect by coating the window with a film one-quarter of a wavelength thick. An appropriate wavelength is 550 nm, which is in the mid-band of the visible spectrum. Under these conditions, the internally and externally reflected rays interfere destructively reducing the reflected intensity to a fraction of 1% (Fig. 3.28). It is shown in any good textbook on optics that, following from (2.20), to satisfy the interference conditions for normally incident light, the refractive index of the film, n_f , should equal $\sqrt{n_g}$, where n_g is the refractive index for glass. There are no suitable coatings with n_f exactly equal to 1.22 although magnesium fluoride, with $n_f = 1.38$, has the desired physical properties and is effective in reducing reflections to less than 1% over the visible spectrum.

This treatment, known as blooming, achieves a form of graded refractive index. An intriguing example of man mimicking nature is discussed by Clapham and Hutley (1973) and later by Wilson and Hutley (1982) concerning the corneal lenses of certain nocturnal moths. The outer eye surface is covered with a regular array of conical protuberances that suppress reflections, thereby concealing the moth's presence. Electron microscope studies reveal that the cones have a height h , and spacing d , both of the order of 200 nm, and analysis shows that, if $h/\lambda > 0.5$, the reflectance is essentially 0. Reflections are suppressed in this case by providing a graded index of refraction between the air and the cornea; photofabricated arrays made by these authors confirm the elimination of reflections. Another novel technique for surface treatment, described by Fischer (1983), is based on creating etched nuclear tracks by bombardment with 750 keV argon ions. Ion energy is set to produce track lengths of $\lambda/2$. These expensive procedures do not find widespread application with PMTs because reflection at the air–glass interface constitutes a relatively small loss, the gaining of which is unlikely to be cost effective. There is, however, a more significant reflection loss occurring at the glass–photocathode interface.

3.13.2 Sandblasting

PMT manufacturers offer products that incorporate textured external window surfaces. The finish is usually a polished one, although certain applications benefit from a roughened finish on the window, depicted in Fig. 3.29(b).

Liquid scintillation counting of low-energy isotopes, such as tritium, is an example in which maximizing light collection is crucial. This application is based



3

Fig. 3.28. A non-reflecting crystalline film.

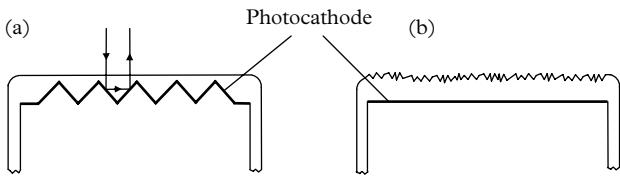


Fig. 3.29. (a) Cross section of a prismatic window illustrating how reflected light has a second interaction with the photocathode. (b) Micro-facets (exaggerated), created by sandblasting, scatter both incoming light and reflected light from the photocathode.

on coincidence counting so that an improvement in individual PMT QEs leads to a twofold gain in detection rate. The benefits of increased QE may be realized in any application where the PMT window is in air or vacuum. A textured surface may be achieved by chemical etching, sandblasting, or abrasion by mechanical means (grinding or sandpapering). There is a considerable literature on sandblasting, with the paper by Bouzid and Bouaouadja (2000) concerning the resilience of soda glass under the conditions of the Sahara desert, of particular relevance to PMTs. These authors have also measured the change in optical transmission (presumably white light) with duration and it appears that the surface roughness achieved is quite insensitive to the sand grain size. For sand, distributed within a size range of 100 to 800 μm , the roughness, R_t , as measured by a surface profilometer, is of the order of 2 to 3 μm , and similarly for the mean depth. Transmission deteriorates with exposure time tending towards 50%, but this loss is at the gain in scattered light. Janecek and Moses (2008a, 2010) have made measurements on surface treatment of BGO crystals obtaining similar results. They express profilometer measurements in terms of a statistical distribution in micro-facet angles, with rms values of 1.3°, 3.8°, and 12° for polished, etched, and ground surfaces, respectively. The corresponding fwhms for scattered light distributions are 6%, 18%, and <58%. This last figure is similar to that quoted by Bouzid and Bouaouadja (2000) and points to the most likely mechanism for optical enhancement.

Randomly textured surfaces, such as those produced by sandblasting, enhance QE, but not in accordance with the graded-index argument because, unlike moth-eye conical distributions, there is no array regularity. Furthermore, the feature size and depth created by sandblasting are of the order of several microns, compared with one-quarter of a micron for an array—diffraction therefore does not occur. It is proposed that the observed enhancement has two components. Normally incident light is refracted by a roughened outer window, imparting an angular distribution. The effect of this is equivalent to an increase in QE because of the generation of longer path lengths in the photocathode. Second, the photocathode has a reflection coefficient of ~20% but, for an optically polished outer window surface, most of this reflected light is transmitted into the air and lost. However, if the window is sandblasted, from the results of studies mentioned, up to half the

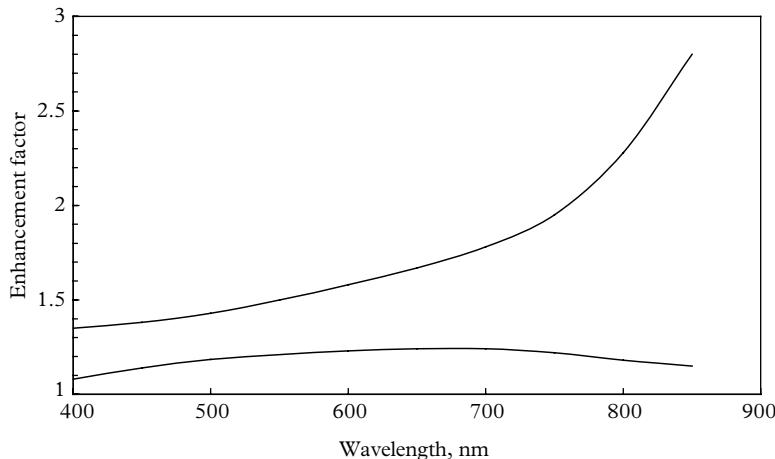


Fig. 3.30. QE enhancement factor for a 2" S20 photocathode after sandblasting (lower curve). The upper curve refers to a prismatic window variant, the 9658B.

reflected light from the cathode is scattered by the diffuse outer window surface, and available for detection.

There are few experimental results on the effects of optical enhancement created by sandblasting. The experience at ET Enterprises is that the degree of enhancement is not critically dependent on the average particle size of material used for ablation, also noted by Bouzid and Bouaouadja (2000). Once a uniform surface roughness has been attained, continuing bombardment has no effect other than to abrade the surface. It is also found that chemical etching produces the same degree of enhancement as does sandblasting. It appears that bialkali PMTs with thin bowed windows, such as the 9829B, or a 9354B hemispherical PMT, provide enhancement of about 20% at 420 nm but the photocathodes of the more common PMTs with plano-planar windows are enhanced by about 10%—the reason for the lowered enhancement is unclear. The plano-planar S20, Type 9558B, for example, shows enhancement over a wide band from 400 to 850 nm, as is evident in Fig. 3.30.

Sandblasting provides an effective key for the inclusion of a wavelength shifter on a PMT window. The enhanced robustness so gained is vital in low-temperature applications, such as operation in liquid argon.

3.13.3 Theory of light recycling

The outcome of light incident on the window of a PMT is as follows: a small fraction of light is reflected at the air–glass interface while the remainder interacts with the photocathode. Some of these photons are absorbed by the photocathode, others are transmitted through the layer without interaction, and the remainder are

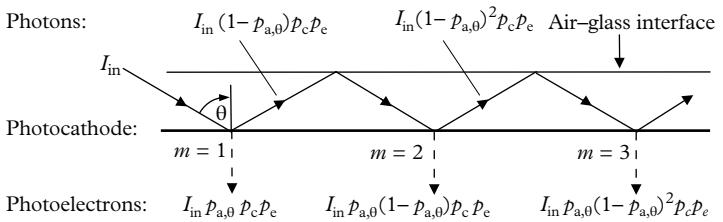


Fig. 3.31. The passage of trapped light in a waveguide formed by the photocathode and the external window surface of a PMT.

reflected. The theory developed in Chapter 2 indicates that the proportion reflected is substantial for red light and the function of enhancers is to return otherwise escaping light to the photocathode. Enhancers also reduce the transmitted component by directing incoming light obliquely to the photocathode. Optical enhancers operate by trapping incident light in the bulk of a PMT window. A light signal is confined to propagation within the external plane of the window and the internal layer of the photocathode; this requires that the launch angle is greater than the critical one. For geometrical considerations, it is customary to choose 45° incidence, which is close to the TIR mode for glass. There is a finite probability of adsorption for a photon at the first encounter with the photocathode surface. If the photon fails to convert, it may do so on a subsequent bounce, and so on, as shown in Fig. 3.31. This is referred to as recycling. Some ingenious schemes are discussed in 3.13.4 to 3.13.6, but first we examine the theory of recycling.

The photoelectric effect, according to the description given by Sommer (1980), is a three-step process. The first is the absorption of an incident photon with probability $p_{a,0}$, leading to the production of an electron in the photocathode layer; the second, p_c , is the probability that the absorbed energy of the photon promotes an electron into the conduction band; and the third, p_e , is the probability that this electron escapes the photocathode into vacuum. Following the approach of Hallensleben *et al.* (2000) for a photon at normal incidence, the QE, η_0 , may be written as

$$\eta_0 = p_{a,0} p_c p_e. \quad (3.12)$$

For light incident at an angle, θ , the QE is $\eta_\theta = p_{a,0} p_c p_e$, on the assumption that p_c and p_e are independent of θ ; η_0 is the parameter quoted by manufacturers, while η_θ is readily measured in an equipment similar to that shown in Fig. 2.10(a).

Consider a narrow parallel beam of light traversing the window of a PMT. The light is launched into the window at an angle greater than the critical one for glass. At the first point of reflection, indicated by $m = 1$, a fraction of the input light, $I_{in}p_{a,0}p_c p_e$, is converted to photoelectrons, and the light remaining, $I_{in}(1 - p_{a,0})p_c p_e$, proceeds to the next reflection point, $m = 2$. At this point, a further $I_{in}p_{a,0}(1 - p_{a,0})^2 p_c p_e$

$p_{a,0})p_c p_e$ photoelectrons are produced, with $I_{in}(1 - p_{a,0})^2 p_c p_e$ photons remaining, and so on. The number of photoelectrons produced after m reflections, I_m , is given by the summation of terms in the photoelectrons line in Fig. 3.31:

$$\begin{aligned} \sum_1^m I_m &= I_{in} p_{a,0} p_c p_e [1 + (1 - p_{a,0}) + (1 - p_{a,0})^2 + \cdots + (1 - p_{a,0})^{m-1}] \\ &= I_{in} p_c p_e [1 - (1 - p_{a,0})^m]. \end{aligned} \quad (3.13)$$

For m sufficiently large, the expression in square brackets tends to unity, and

$$\sum_{m=1}^{\infty} \frac{I_m}{I_{in}} = p_c p_e.$$

represents the total number of photoelectrons released in a multiple bounce arrangement, where the summation on the left hand side is the corresponding QE, $\eta_{\Sigma,0}$. Recalling that $\eta_0 = p_{a,0} p_c p_e$ is the QE at normal incidence quoted by manufacturers, we have

$$\eta_{\Sigma,0} = p_c p_e = \eta_0 / p_{a,0}, \quad (3.14)$$

with a QE enhancement factor of $1/p_{a,0}$. Since $p_{a,0}$ is always less than unity, we can expect enhancement with this arrangement for all wavelengths, and for all photocathode types. The simple relationship in (3.14) may at first appear remarkable but after some thought it becomes obvious that it must be of this form with the enhancement factor varying as $1/p_{a,0}$. Note the similarity in approach to that adopted for an integrating sphere leading to (3.5). Equation (3.14) predicts that the potential enhancement at wavelengths for which $p_{a,0}$ is small, is significant. This is verified experimentally in 3.13.5: the enhancement factor for blue light is in the region of 1.2 to 1.5 times, extending to an enhancement of up to 10 at infrared wavelengths. The achievable enhancement also depends on the type of photocathode (e.g. whether it is a bialkali or an S20 photocathode).

3.13.4 Internal prismatic window

The earliest commercially available PMT incorporating an optical enhancer was the 9658B produced by EMI in the early 1970s. Enhancement is produced by an internal pattern of four-sided pyramids of height d , and spacing h , which serve as the photocathode substrate. This is illustrated in Fig. 3.29(a), which shows how light normally incident on the PMT window strikes one of the photocathode facets. A fraction of the incident light is reflected and then intercepted by a neighbouring facet, with the production of further photoelectrons. There are thus two contributions to enhancement with this arrangement: one from the longer paths in the photocathode material, and one from capturing reflected light from the first photocathode impact. It is evident in Fig. 3.30 that enhancement occurs at

all wavelengths but it is most effective in the infrared, where the photocathode is relatively transparent. The extraction field gradient at the photocathode is weakened by the presence of pyramids and consequently some photoelectrons are not collected but return to the photocathode. This explains why increasing the height of the pyramids may yield more photoelectrons, but a higher proportion of these are trapped, with no net benefit. Note the similarity between photoelectron collection and secondary emission in Venetian blind dynodes (see 5.6.1). One way to determine the enhancement factor of a prismatic PMT is to compare test parameters of a sufficiently large sample of such PMTs with the plano-planar version of the same type. Well-known production variability disfavours this approach but special devices devised by EMI have overcome this problem through combining the two photocathode geometries within the same envelope: half the window area includes the prism pattern, while the other half remains plane. Determining the enhancement factor with sandblasted plano-planar windows does not need special devices, as measuring the QE before and after sandblasting also provides the answer.

3.13.5 External optical enhancement devices

Numerous optical devices have been devised to enhance QE based on creating zigzag light paths within the window of an end window PMT, illustrated in Fig. 3.32. The optics for launching light is the critical element in any optical enhancer. Gunter *et al.* (1970) review most of the variants in a comprehensive investigation. Following the first application by Rambo (1964), several variants have been published by W D Gunter and various colleagues at the Ames Research Center, Mountain View, California, USA, beginning in 1965. There are essentially

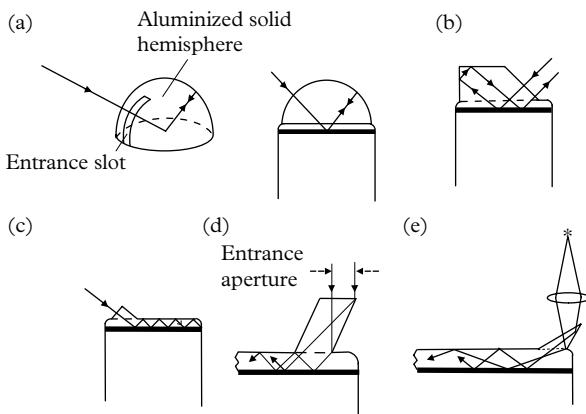


Fig. 3.32. (a, b) Two-bounce enhancers due to (a) Oke and Schild (1968) and (b) Gunter *et al.* (1970). (c–e) Multi-bounce enhancers utilizing various prism arrangements for launching.

two groups of enhancers: the first comprises the two-bounce enhancers, and the second comprises enhancers with multiple reflections. Fig. 3.32 (a) and (b) show the two-bounce variety. In Fig. 3.32(a), light enters through a narrow window in the otherwise aluminized solid quartz hemisphere; it is either reflected by the photocathode and re-imaged after reflection by the aluminizing, or it is absorbed. The configuration in Fig. 3.32(b) relies on TIR and is therefore bare. Measurements by Oke and Schild (1968) based on the ITT FW130 PMTs are shown in Fig. 3.33, together with results from Gunter *et al.* (1970). The performance of an EMI 9658B, which has an internal two-bounce enhancer, is also shown in Fig. 3.33. The general nature of the response is the same in all three cases, but the degree of enhancement will depend on the sensitivity of the selected PMT. Oke and Schild (1968) present results for seven S20, ITT FW130 PMTs that show a range of almost a factor of 2 in enhancement.

Multi-bounce arrangements similar to those depicted in Fig. 3.32(c)–(e) offer the maximum attainable performance, especially in the infrared. The 90° prism in Fig. 3.32(c) may be replaced with a glass quadrant giving similar performance. The disadvantage of mandatory off-axis and oblique incidence in Fig. 3.32(c) may be removed by the schemes shown in Fig. 3.32 (d) and (e); the prism in both cases can be located on the optic axis of a PMT by translation, but the maximum number of bounces is halved. The choice of a 60° prism in the enhancer of Hallensleben *et al.* (2000) increases the propagation angle in the window, approaching the ideal of 75°, for which the theory developed by these authors predicts optimum performance.

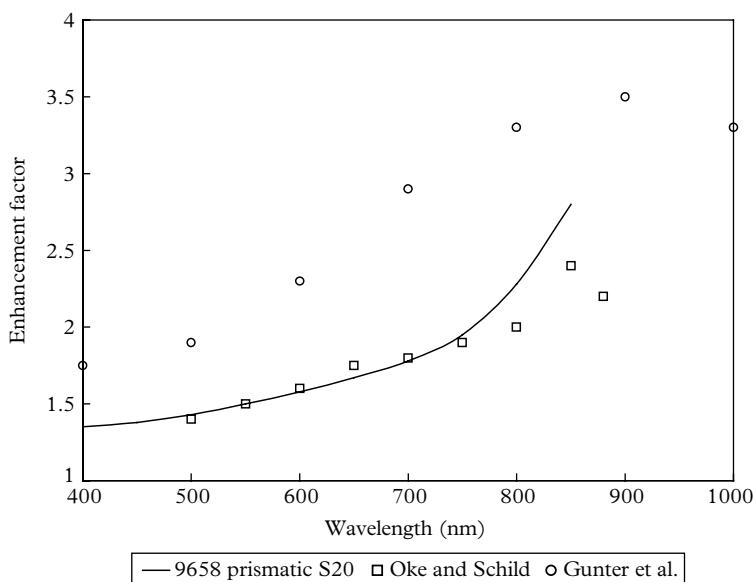


Fig. 3.33. Comparison of two-bounce enhancers with a prismatic EMI 9658B.

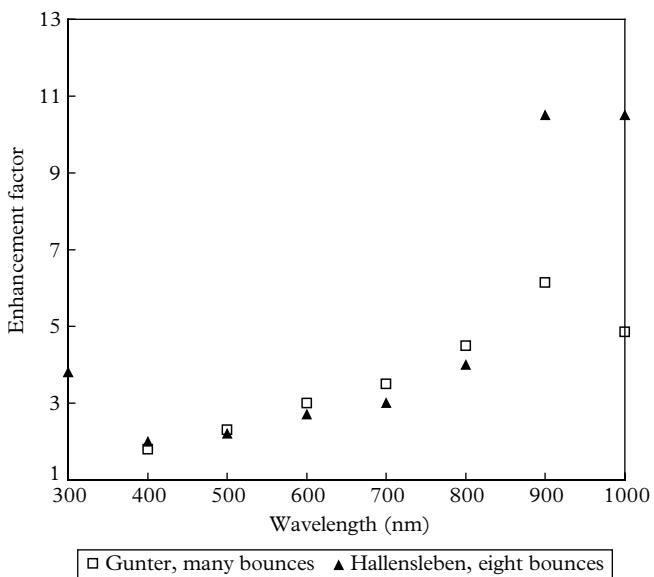


Fig. 3.34. Multiple bounce performance.

Gunter, W. D., Grant, G. R. and Shaw, S. A. (1970). Optical devices to increase photocathode quantum efficiency. *Appl. Opt.*, 9, No. 2, 251–7. Reprinted with permission from OSA ©1970.

Results from Gunter *et al.* (1970) and Hallensleben *et al.* (2000), reproduced in Fig. 3.34, agree over the visible range of the spectrum. For wavelengths falling outside this region, we note exceptional performance in the results of Hallesleben *et al.* The four-fold enhancement in the ultraviolet needs independent validation since it implies a QE in the region of 50%. These authors do not quote the state of polarization, which is important since enhancement is sensitive to this parameter; aligning optics of this type of enhancer is known to be difficult and frustrating—especially where the light is invisible, and this may have affected measurements by these authors. By comparison, S Shaw *et al.* (1971), using the arrangement shown in Fig. 3.32(d), measured an enhancement ratio of 1.5 for wavelengths between 230 and 400 nm, based on a 9558QB PMT.

In summary, the use of optical enhancers applies to collimated light sources with beam widths of a few millimetres only. Dark count is a limiting factor in achieving high signal-to-background performance, although cooling an S20 photocathode typically reduces this parameter by up to 100 times. Unfortunately, the infrared response of all PMTs decreases significantly with cooling and there is thus an optimum operating temperature for every PMT to achieve the best signal-to-background ratio. Dark counts increase as the square of the PMT diameter (although there are other considerations) and hence it is always advisable to choose a PMT with a diameter to match the geometry of the source. It should be obvious from Fig. 3.32 that window thickness is part of the design and PMTs

with a plano-concave window are unsuitable. A common limitation to all the published enhancement schemes is the requirement for collimated light with a maximum acceptance angle of about 5° .

3.13.6 WLS coatings

WLS compounds absorb UV light and re-admit isotropically at visible wavelengths, where PMTs have their maximum QE. The method, shown in Fig. 3.35, requires a film thickness of only a few microns, and yet conversion can approach 100%.

Windows of 2 mm thickness cut off at the following wavelengths: standard borosilicate, 300 nm; UV glass, 185 nm; and fused silica (quartz), 160 nm. The WLS technique is eminently suitable for the three window options in applications requiring sensitivity in the wavelength region 90 to 300 nm. This technique provides the opportunity of replacing a quartz window PMT with a less expensive one of standard borosilicate glass, although the composite device provides only half the QE in the 200 to 300 nm band, compared with a standard quartz PMT. This is because half the converted light is directed away from the photocathode. Light of longer wavelength than 350 nm is unaffected, and hence it is detected with the same efficiency as for an untreated PMT. The decay time of the binder and the WLS is important where fast timing is required. Eigen and Lorenz (1979) report decay times of 1 ns and 2.5 ns, with a low-intensity slow component of 3.5 ns, depending on the choice of binder; most WLS dyes appear to have fast decay characteristics.

Sodium salicylate and p-terphenyl were among the earliest of the selected compounds and the paper of Nygaard (1965) provides useful reference to the

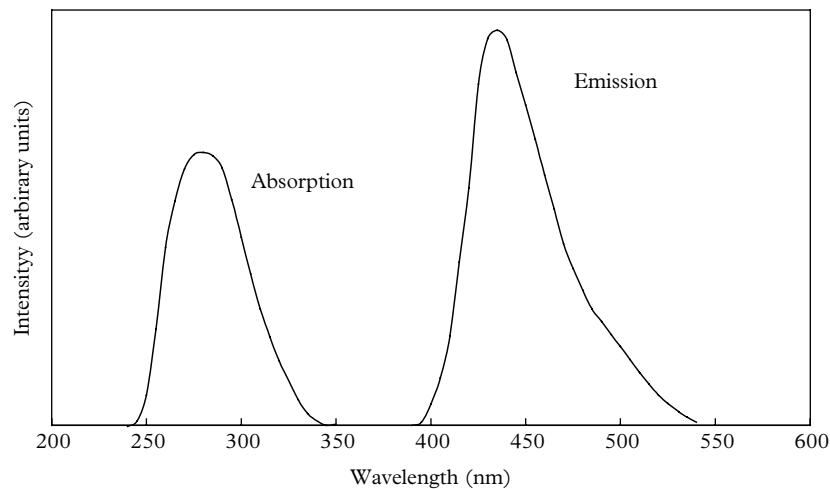


Fig. 3.35. Typical absorption and emission spectra for a wavelength shifter dye.

activity at that time. These materials are now seldom used: sodium salicylate is hygroscopic and discolours under bright light. Furthermore, films produced by vacuum deposition on to a glass slide, or directly on to the window of a PMT, tend to be mechanically fragile. Eigen and Lorenz (1979) describe the inclusion of a transparent binder with the WLS material and a solvent. Their technique consisted of dipping the front end of a PMT into the mixture, removing it, and allowing the solvent to evaporate. The process is quick and the final product sufficiently durable. The WLS material selected was PTP (1,4-diphenylbenzol ($C_{18}H_{14}$)), a binder of Paraloid B72, dissolved in methylene dichloride (CH_2Cl_2). Details of the proportions used, together with other practical information may be found in their publication. A subsequent publication by Saito *et al.* (2009) provides yet further options. Viehmann and Frost (1979) used mixtures of WLS materials, together with Elvacite 2045 binder dissolved in toluene. These authors give performance information for a range of materials.

One of the outstanding scintillation detectors ever built is the massive neutrino detector ICARUS, which deploys hemispherical PMTs (9357FLA). There are two formidable technical difficulties: first, PMTs are immersed in liquid argon (84 K), and second, the scintillation light is centred on 128 nm. Most scintillators are transparent to their own light so that WLS is most effective at the PMT window. The chosen WLS material is tetraphenyl butadiene (TPB), which down-converts to a peak emission at 430 nm, spanning 390 to 520 nm—an excellent match to the bialkali QE response curve. The authors claim 100% fluorescence efficiency and 2.5 times greater light output than sodium salicylate. The PMT window is sandblasted on the outside to provide a key for the TPB; a semi-transparent platinum substrate supporting the photocathode mitigates its otherwise high resistivity at low temperature. The overall QE is 8 to 11%, a significant achievement considering the converted photon emission is isotropic, and therefore half is lost anyway.

To summarize, the QE of a given photocathode depends primarily upon the angle of incident of the light, its wavelength, and polarization. Strictly speaking, the optical arrangements so far mentioned increase photosensitivity without changing QE: the multi-bounce schemes described in Fig. 3.32 increase photoelectron emission by means of photon recycling; similarly, WLS coatings transform incident light to longer wavelengths where QE is higher. Such methods do not strictly increase QE, but equate to an equivalent increase in QE, through enhanced light gathering. Most authors do not make this distinction, but their meaning is usually clear.

REFERENCES

- Albrow, M. *et al.* (1999). A preshower detector for the CDF Plug Upgrade. *Nucl. Instr. and Meth. in Phys. Res. A*, **431**, 104–11.
- D'Ambrosio, C. *et al.* (1999). Improved light transmissions from scintillators to new photocathode windows. *Nucl. Instr. and Meth. in Phys. Res. A*, **431**, 455–9.

- Aota, S. *et al.* (1999). Mass production of tile/fibre units for the CDF plug upgrade EM calorimeter. *Nucl. Instr. and Meth. in Phys. Res. A*, **420**, 48–61.
- Artikov, A. *et al.* (2006). New generation large area muon scintillation counters with wavelength shifter fibre readout for CDF II. *Phys. Part. Nuclei*, **3**, 188–200.
- Asakawa, T. *et al.* (1994). A scintillating tile/fibre system for high energy calorimeter with high light yield and uniform longitudinal and transverse response. *Nucl. Instr. and Meth. in Phys. Res. A*, **340**, 458–65.
- Barish, B. *et al.* (1978). Very large area scintillation counters for hadron calorimetry. *IEEE Trans. NS*, **25**, 532–6.
- Barnaby, C. F. and Barton, J. C. (1960). Performance of large area scintillation counters. *Proc. Phys. Soc.*, **76**, 745–53.
- Barton, J. C., Crispin, A., and Slade, M. (1964). The efficiency and transparency of cheap liquid scintillators. *J. Sci. Instr.*, **41**, 736–9.
- Bouzid, S. and Bouaouadja, N. (2000). Effect of impact angle on glass surfaces eroded by sand blasting. *J. Eur. Ceram. Soc.*, **20**, 481–8.
- Chen, J., Zhang, L. and Zhu, R. (2005). Large size LYSO crystals for future high energy physics experiments. *IEEE Trans. NS*, **52**, 3133–40.
- Clapham, P. B. and Hutley, M. C. (1973). Reduction of lens reflection by the “moth eye” principle. *Nature*, **224**, 281–2.
- Eigen, G. and Lorenz, E. (1979). A method of coating photomultipliers with wavelength shifters. *Nucl. Instr. and Meth.*, **166**, 165–8.
- Evans, R. D. (1955). *The Atomic Nucleus*. McGraw-Hill, New York.
- Fraser-Mitchell, J. and Wright, A. G. (1990). Contribution of photocathode nonuniformity to energy resolution in NaI(Tl) scintillation detectors. *Nucl. Instr. and Meth. in Phys. Res. A*, **228**, 429–38.
- Fukui, S. *et al.* (1971). Flexible adiabatic light guide of silicon rubber. *Jap. J. of Appl. Phys.*, **10**, 1597–600.
- Fischer, B. E. (1983). A new antireflection treatment for optical surfaces by the nuclear track technique. *Nucl. Instr. and Meth.*, **216**, 265–8.
- Hinterberger, H. and Winston, R. (1966). Efficient light coupler for threshold Cerenkov counters. *J. Sci. Instr.*, **37**, 1094–5.
- Hinterberger, H. and Winston, R. (1968). Use of a solid light funnel to increase phototube aperture without restricting angular acceptance. *J. Sci. Instr.*, **39**, 1217–18.
- Garwin, R. L. (1952). The design of liquid scintillation cells. *Rev. Sci. Instr.*, **23**, 755–7.
- Gunter, W. D. Grant, G. R. and Shaw, S. A. (1970). Optical devices to increase photocathode quantum efficiency. *Appl. Opt.*, **9**, 251–7.
- Hallensleben, S., Harmer, S. W., and Townsend, P. D. (2000). Optical constants for the S20 photocathode, and their application to increasing photomultiplier quantum efficiency. *Opt. Commun.*, **180**, 89–102.
- Janecek, M. (2012). Reflectivity spectra for commonly used reflectors. *IEEE Trans. NS*, **59**, 490–6.

- Janecek, M. and Moses, W. W. (2008a). Optical reflectance measurements for commonly used reflectors. *IEEE Trans. NS*, **55**, 2432–7.
- Janecek, M. and Moses, W. W. (2008b). Measuring light reflectances of BGO crystal surfaces. *IEEE Trans. NS*, **55**, 2443–9.
- Janecek, M. and Moses, W. W. (2010). Simulating scintillator light collection using measured optical reflectance. *IEEE Trans. NS*, **57**, 964–70.
- Jordan, D. V. *et al.* (2003). *Progress Report on the Advanced Large-Area Plastic Scintillators (ALPS) Project*. Prepared for the US Department of Energy under contract DE-AC06-76RL01830. Pacific Northwest National Laboratory, Richland, Washington.
- Keil, G. (1970). Design principles of fluorescence radiation converters. *Nucl. Instr. and Meth.*, **87**, 111–23.
- Koechlin, Y. (1955). Caractéristiques du montage optique d'un scintillateur. *J. de Phys. et le Rad.*, **16**, 849–53.
- Labsphere. (2012). *Technical Guide: Integrating Sphere Radiometry and Photometry*. Labsphere, North Sutton, NH. Available at <http://www.labsphere.com/site/assets/files/2550/a-guide-to-integrating-sphere-radiometry-and-photometry.pdf> (accessed 20 Jan 2017).
- Lovell, D. J. (1984). Theory and applications of integrating sphere technology. *LFW*, **20**, 86–96.
- McMillan, J. E. (2005). Improvements to light collection in scintillation detectors intended for cosmic ray timing. *29th International Cosmic Ray Conference, Pune, India*, pp. 283–6.
- Moorhead, M. E. and Tanner, N. W. (1996a). Optical properties of an EMI K₂CsSb bialkali photocathode. *Nucl. Instr. and Meth. in Phys. Res. A*, **378**, 162–70.
- Moorhead, M. E. and Tanner, N. W. (1996b). Light reflecting concentrators for photomultipliers with curved photocathodes. *Appl. Opt.*, **35**, 3478–86.
- Oke, J. B. and Schild, R. E. (1968). A practical multiple reflection technique for improving the quantum efficiency of photomultiplier tubes. *Appl. Opt.*, **7**, 617–22.
- Nygaard, K. (1965). Instrumental problems related to the measurement of quantum efficiency of sodium salicylate. *J. Opt. Soc. Am.*, **55**, 944–6.
- Ostankov, A. *et al.* (2000). A study of the new hemispherical 6-dynodes PMT from Electron Tubes. *Nucl. Instr. and Meth. in Phys. Res. A*, **442**, 117–23.
- Rambo, B. E. (1964). *Improved Long Wavelength Response of Photoemissive Surfaces*. Air Force Technical Report ALTDR 64–19. Air Force Avionics Lab, Wright-Patterson AFB, Ohio.
- Saito, T. Y. *et al.* (2009). Very high QE HPDs with a GaAsP photocathode for the MAGIC telescope project. *Nucl. Instr. and Meth. in Phys. Res. A*, **610**, 258–61.
- Saint-Gobain. (2016). *Scintillating Optical Fibres*. Saint-Gobain Ceramics & Plastics, Inc., Paris.
- Kuraray. (n.d.) *Plastic Scintillating Fibers (Materials and Structures)*. Kuraray Co., Ltd., Tokyo. Available at <http://kuraraypsf.jp/psf/> (accessed 20 Jan 2017).

- Shaw, P.-S. *et al.* (2007). Ultraviolet characterisation of integrating spheres. *Appl. Opt.*, **46**, 5119–28.
- Shaw, S. A., Grant, G. R., and Gunter, W. D. (1971). Optical enhancement of photomultipliers at ultraviolet wavelengths. *Appl. Opt.*, **10**, 2559.
- Shurcliff, W. A. (1950). Radiance amplification by multi-stage fluorescence system. *J. Opt. Soc. Am.*, **41**, 209.
- Sommer, A. H. (1980). *Photoemissive Materials*. Robert E Krieger Publishing Company, New York.
- Springsteen, A. (2009). *Integrating Sphere Coatings: A Historical Perspective*. Avian Technologies LLC, Sunapee, NH. Available at http://cormusa.org/uploads/CORM_2009_SphCoat_Springsteen.pdf (accessed 19 Jan 2017).
- Swank, R. K. (1954). Characteristics of scintillators. *Ann. Rev. Nucl. Sci.*, **4**, 111–40.
- Viehmann, W. and Frost, R. L. (1979). Thin film wavelength shifter coatings for fluorescent radiation converters. *Nucl. Instr. and Meth. in Phys. Res. A*, **167**, 405–15.
- Vilardi, I. *et al.* (2006). Optimization of the effective light attenuation length of YAP:Ce and LYSO:Ce crystals for a novel geometrical PET concept. *Nucl. Instr. and Meth. in Phys. Res. A*, **564**, 506–14.
- Vittori, F. *et al.* (2000). A study on light collection of small scintillating crystals. *Nucl. Instr. and Meth. in Phys. Res. A*, **452**, 245–51.
- Welford, W. T. and Winston, R. (1978). *The Optics of Nonimaging Concentrators: Light and Solar Energy*. Academic Press, New York.
- Wilson, S. J. and Hutley, M. C. (1982). The optical properties of ‘moth eye’ antireflection surfaces. *Opt. Acta*, **29**, 993–1009.
- Wigmans, R. (2000). *Calorimetry: Energy Measurement in Particle Physics*. Oxford Science Publications. Oxford University Press, Oxford.
- Zhu, R. Y. *et al.* (1996). A study on the properties of lead tungstate crystals. *Nucl. Instr. and Meth. in Phys. Res. A*, **376**, 319–34.

4

Statistical processes

4.1 Introduction

Two physical processes are central to the operation of all PMTs: the photoelectric effect describes the conversion of incident photons to photoelectrons, while the electron multiplier provides gain. Both of these processes are statistical in nature. The binomial distribution (also known as the Bernoulli distribution) describes photoelectron statistics—on the basis that either a photoelectron is produced or it is not. The Poisson distribution is a special case of the binomial distribution, as is the normal (Gaussian) distribution. PMT operation by its very nature is highly statistical. Familiarity with these statistical processes undoubtedly generates an ability to exploit the full potential offered by PMTs. The author has found the statistical coverage of the books by the following authors to be of considerable value: Evans (1955), Jánossy (1965), and Riley *et al.* (2002). Statistical formulations and the predictions that follow are developed mostly from first principles. Generating functions are used extensively for this purpose, as they are both elegant and capable of providing answers more efficiently than traditional approaches that lead to formulations for mean, variance, and higher-order moments. Three distributions—binomial, Poisson, and, to a lesser extent, normal—play important roles in describing the way in which PMTs operate; statistics are also fundamental to a wide range of applications involving pulse height resolution and timing. This chapter presents a theoretical foundation, based on statistical arguments, to explain some of the fundamental processes occurring within a PMT. A combined didactic and phenomenological treatment for topics with a high degree of statistical argument invariably results in some overlap with coverage in other chapters. For example, the statistical nature of noise, resolution, timing, and afterpulses is covered in this chapter in addition to an experimental treatment in other sections, where appropriate.

Perhaps the most critical of all the processes within a PMT is that of photo-emission, for this is where the conversion of a light signal into an electrical one takes place. From a signal recovery point of view, it is obvious that, the higher the number of photons converted by the photocathode, the greater is its sensitivity—specifically, the ability to detect feeble light signals. This process is quantified by assigning a number, the QE, indicating how good, or bad, the result is likely to be. Likewise, secondary electron emission—the process that provides gain—is statistical and therefore noisy. PMTs are capable of precise time-stamping: that is, the ability to relate the time at which an anode signal occurs to the time of its generation, with sub-nanosecond precision. PMTs are also capable of resolving precise time intervals between events.

4.1.1 Single- and multi-photon light sources

Detection and multiplication processes in PMTs are statistical in character, as previously discussed. Light sources relevant to PMTs are generally of this nature

and it is therefore appropriate to discuss the requirements for creating steady or pulsed sources for PMT testing. Photon-counting instrumentation is based on the detection of single photons: equipment such as alpha probes and luminometers are well-known examples. Alternatively, scintillation spectrometers, for example, detect multi-photon signals, and test equipment requires a pulsed LED or laser, or a scintillator. General test and selection procedures, for both manufacturers and users, require single- and multi-photon sources and it turns out that devising such sources is straightforward. Consider a photocathode weakly illuminated with single photons, producing a low rate of photoelectrons of n_e per second. If the mean gain of a PMT is $\langle g \rangle$ the anode current is

$$I_e = n_e e \langle g \rangle.$$

Increasing the light input to produce multi-photoelectron pulses of average size μ and at a rate of n_μ per second generates an anode current:

$$I_\mu = \mu n_\mu e \langle g \rangle.$$

The ratio of the two currents leads to μ , noting that knowledge of $\langle g \rangle$ is unnecessary in this consideration, but it must remain fixed in the transition from a single to a multi-photon source:

$$\mu = \frac{n_e I_\mu}{n_\mu I_e}. \quad (4.1)$$

A pulsed single-photon source of sufficient purity can be derived from a DC light source by ensuring that the probability of recording two pulses, $p(2)$, per sample period is vanishingly small. A simple test with an MCA will readily confirm whether a source is single photon: enhancing the brightness should only increase the integral count rate without changing the shape of the SER, and, if there is one, the position of the distribution peak. The set-up procedure can be performed with an oscilloscope, but not so conveniently. The same considerations apply to a pulsed light source that has the additional advantage of synchronicity. Following from (4.2), for a light source obeying Poisson statistics, the ratio of the probability of recording two photons to that for single photons is $p(2)/p(1)$, in unit time:

$$\frac{p(2)}{p(1)} = \frac{e^{-\mu} \frac{\mu^2}{2!}}{e^{-\mu} \frac{\mu^1}{1!}} = \frac{\mu}{2}. \quad (4.2)$$

Statistically, we require an ‘inefficient’ source, with $p(2)/p(1)$ as small as practical, which corresponds to setting the mean, μ , at $\sim 10^{-3}$; this reduces the two-photon contribution to below $\sim 0.05\%$, and the probability of generating three or more photons is negligible. Note that the multi-photoelectron level, $\eta\mu$, is always lower than μ . Consequently, for such a pulsed light source, the most likely outcome is the absence of a photoelectron output, with $p(0)$ very close to unity. A current-pulsed LED is inexpensive and convenient for generating light pulses with a width

ranging upwards from ~ 2 ns. Fast circuits are generally based on an avalanche transistor, details of which are given in Chapter 14, Fig. 14.28. Photons incident on a photocathode should be random, free from correlations, and incident at a low mean rate: that is, spaced out. However, light sources such as LEDs are unstable unless sufficiently energized. This requirement may be satisfied, for example, by using the combination of a relatively bright light source (by PMT standards) and sufficient neutral density filters.

A single-photon source can also be derived from a DC light source. If a LED is chosen, the activating current must be sufficient to avoid the generation of bursts, and, as in the foregoing discussion, the mean time interval between photoelectrons must be sufficient to minimize bunching (see 4.18.1). The advantage of a pulsed light source over a DC one lies in its synchronicity. A light source configured to generate a multiplicity of photoelectrons per pulse finds application in evaluation of basic PMT performance.

4.2 Binomial, Poisson, and normal distributions

The symbolic representation of probabilities follows the convention of separating variables and parameters with a semi-colon. For example, $P(n; \mu, \sigma)$ means the probability of observing n in a distribution with parameters μ and σ . Consider the detection of a finite number of photons, N , under the assumption that the behaviour (or fate) of individual photons is completely independent. The probability for detection of each photon is p , and the probability for failing to detect is therefore $q = (1 - p)$. The probability that a particular n is detected and the remaining $(N - n)$ not detected is $p^n(1 - p)^{N-n}$. The probability that any n will be detected exceeds $p^n(1 - p)^{N-n}$ by the factor ${}^N C_n$, which is the number of ways of selecting n out of N . Combining these expressions leads to the binomial distribution

$$P(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}. \quad (4.3)$$

Although $P(n; N, p)$ customarily refers to N small, it applies exactly even when N is large. The binomial distribution is so named because $P(n; N, p)$ is identical to the expansion of a power of a sum:

$$(p + q)^N = \sum_{n=0}^N {}^N C_n p^n q^{N-n}.$$

Considering N large and p very small, which applies to most radioactive work and low-light-level detection, the mean detection rate is $r = Np$, and (4.3) becomes

$$\begin{aligned}
P(n; N, p) &= \frac{N!}{n!(N-n)!} \left(\frac{r}{N}\right)^n \left(1 - \frac{r}{N}\right)^{N-n} \\
&= \frac{N^n}{n!} \frac{r^n}{N^n} \left(1 - \frac{r}{N}\right)^N; \\
P(n; r) &= \frac{r^n}{n!} e^{-r}.
\end{aligned}$$

4

This is the Poisson distribution, in which, following convention, r is replaced by the symbol μ —the mean of the distribution—leading to

$$P(n; \mu) = \frac{\mu^n}{n!} e^{-\mu}. \quad (4.4)$$

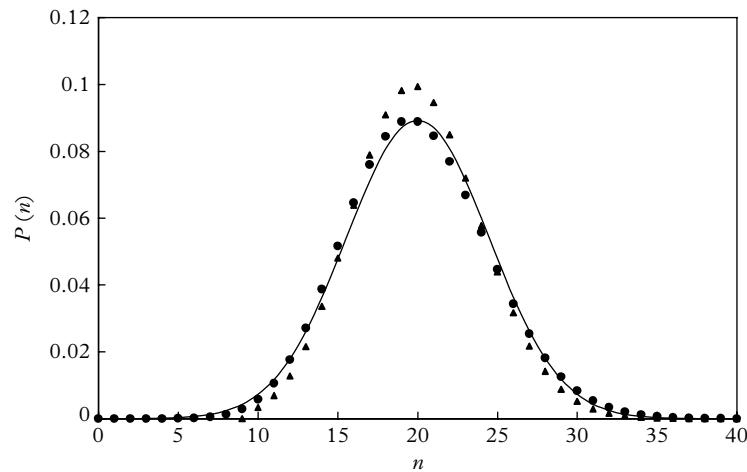
The binomial and Poisson distributions (4.3) and (4.4) present computational difficulties arising from the magnitude of $N!$ and $n!$ when r is large—say, greater than 100. This limitation is overcome by transforming (4.4) using the Stirling approximation, to arrive at the normal distribution, $P(n; \mu, \sigma)$. The manipulations leading to the normal distribution, (4.5) are complicated and tiresome, and given by Jánossy (1965):

$$P(n; \mu, \sigma) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left(-\frac{(n-\mu)^2}{2\sigma^2}\right). \quad (4.5)$$

Using similar arguments, it can be shown theoretically that the binomial also tends to a normal distribution when N is large. Note that the normal distribution is continuous in n , whereas the binomial and Poisson distributions only allow integers. Various measures of dispersion, for both theoretical and experimental distributions, are used: principally, standard deviation, σ , and variance, σ^2 , denoted by ‘var’. Figure 4.1 illustrates the convergence of the three statistical formulations, when n is sufficiently large. For the binomial distribution, N was taken as 100, beyond which computation increasingly lies outside the range of personal computers. The variance of a Poisson distribution is identical to the mean, μ , whereas, for a normal distribution, σ may be chosen at will.

4.3 Mean and variance

The mean can be estimated from a finite set of measurements, although the exact value of μ cannot be determined. This is because μ relates to an infinite population of possible readings. If we have n measurements then we can estimate the mean, denoted by ω , where



4

Fig. 4.1. Three distributions with a common mean: $P(n; N, p)$, a binomial distribution where $N = 100$, and $p = 0.2$ (\blacktriangle); $P(n; \mu)$, a Poisson distribution, for a mean of 20 (\bullet); and $P(n; \mu, \sigma)$, a normal distribution with the same mean and variance as the Poisson (—). Note how closely the three distributions align.

$$\langle x \rangle = \frac{1}{n} \sum_{i=1}^n x_i \approx \mu.$$

The ‘expectation value’ of any statistical variable, denoted by $\langle x \rangle$, is synonymous with the mean: the ‘expectation value’ of x is thus μ , by definition. That $\langle x \rangle$ is always an estimate for experimental data is obvious if one more measurement is made; a total of $n + 1$ instead of n would inevitably lead to a different estimate for μ but the difference can be made vanishingly small if n is sufficiently large. The standard deviation, σ , and the variance, var , are the accepted measures for describing the spread in a distribution, whether it is experimental or analytical:

$$\sigma^2 = \text{var} = \frac{1}{n} \sum_{i=1}^n (x_i - \langle x \rangle)^2.$$

As stated above, if dealing with a finite set of measurements, μ and σ can never be known exactly. For a finite set of observations, the expression for σ^2 must be modified as follows:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \langle x \rangle)^2.$$

The term $(n - 1)$ acknowledges that n readings were analysed to compute $\langle x \rangle$, with the loss of one degree of freedom, leaving a set of $(n - 1)$ independent data from which to calculate σ . It is obvious in the case where we have only one

observation, x_1 , that $\langle x \rangle = x_1$ and σ is indeterminate. The expression for σ^2 including $(n - 1)$ is known as the ‘sample variance’, highlighting the fact that it refers to a finite set of data. It is not the intention to dwell on these finer points of mathematical statistics but it is considered necessary to draw the reader’s attention to such considerations as a precaution: some authors of articles and books involving statistics do not make clear which definition of σ applies. Another point worth noting is that while σ is the most widespread measure of the dispersion of a set of data, or of a mathematical distribution, it is not the only one. Other computations called ‘moments’ are sometimes used, particularly for skewed distributions; detailed study on these topics is reserved for 4.5. In the following sections, we will be dealing with probability distributions or sets of data where n is large and we are justified in using the variance rather than the sample variance.

Given a set of n numbers, where n is large, the variance is

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2. \quad (4.6)$$

Here X is the sample space under consideration, representing all possible outcomes of an individual trial, whereas x represents a particular outcome. We can manipulate (4.6) to express the variance in a more meaningful way. Expanding, we have

$$\begin{aligned} \text{var}(X) &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{2\mu}{n} \sum_{i=1}^n x_i + \frac{\mu^2}{n} \sum_{i=1}^n 1 \\ &= \langle x^2 \rangle - 2\mu^2 + \mu^2, \end{aligned} \quad (4.7)$$

$$\text{var}(X) = \langle x^2 \rangle - \mu^2 = \langle x^2 \rangle - \langle x \rangle^2;$$

$\text{var}(X)$ is thus the average of x^2 minus the average of x , squared.

In the treatment leading to (4.7), the entries may be generated from a formula or they may be experimental. Where the numbers x relate to a probability distribution function, $p(x)$, it is possible to compute the expected number of ‘successes’. This is called the expectation value of X and is written $\langle x \rangle$ or $E[X]$. Note that physicists are not always consistent in their nomenclature and will use upper and lower case variables arbitrarily. In all, there are four different representations for the mean, all of which are utilized:

$$\langle x \rangle = E[X] = \sum_{x=0}^{\infty} x p(x) = M_1 = \mu, \quad (4.8)$$

where for brevity $\sum_{x=0}^{\infty}$ is sometimes condensed to \sum_x .

The expression $\sum_x x p(x)$ will be familiar as the formula for calculating the mean of a set of readings, and the significance of M_1 is explained in 4.5.1. We note that $E[X(X + 1)] = E[X^2] + E[X]$, since by definition

$$E[X(X + 1)] = \sum_x x(x + 1)p(x) = \sum_x x^2 p(x) + \sum_x x p(x) = E[X^2] + E[X],$$

and, similarly,

$$E[X(X - 1)] = E[X^2] - E[X]. \quad (4.9)$$

A function $f(x)$ has an expectation value given by

$$\langle f \rangle = E[f(X)] = \sum_x f(x)p(x).$$

We can generalize these relationships to apply to continuous distributions, such as the normal distribution, where all values of x are allowed, both positive and negative:

$$\langle x \rangle = E[X] = \int_{-\infty}^{+\infty} x p(x)dx, \quad (4.10)$$

$$\langle f \rangle = E[f(X)] = \int_{-\infty}^{+\infty} f(x)p(x)dx. \quad (4.11)$$

For a probability distribution $p(X)$, the expression for the variance, in the usual nomenclature, is

$$\begin{aligned} V[X] &= E[(X - \mu)^2] \\ &= E[X^2] - \mu^2. \end{aligned} \quad (4.12)$$

Equation (4.12) has the same form as (4.9) and was derived from similar reasoning; $V[X]$ is referred to as σ^2 or $\text{var}(X)$ —the terms are synonymous.

We now consider the Poisson distribution to illustrate these considerations. The probability of detecting x events in a process described by Poisson statistics is

$$p(X) = \frac{e^{-\mu}\mu^x}{x!}$$

and, following from (4.8),

$$E[X] = \sum_{x=0}^{\infty} \frac{x e^{-\mu} \mu^x}{x!}.$$

The $x = 0$ term in the expansion of $E[X]$ is zero and may be dropped so that

$$E[X] = \sum_{x=1}^{\infty} \frac{x e^{-\mu} \mu^x}{x!} = e^{-\mu} \mu \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} = e^{-\mu} \mu e^{\mu} = \mu. \quad (4.13)$$

Of course, we knew this anyway because the mean of the Poisson distribution is μ by definition. However, (4.13) illustrates the methodology of determining $E[X]$ for any probability distribution. Using a similar manipulation as in (4.13), noting that the first two terms of the series are zero, we have

$$E[X(X - 1)] = \sum_{x=0}^{\infty} x(x-1)e^{-\mu} \frac{\mu^x}{x!} = e^{-\mu} \mu^2 \sum_{x=2}^{\infty} \frac{\mu^{x-2}}{(x-2)!} = e^{-\mu} \mu^2 e^{\mu} = \mu^2,$$

from which it follows that

$$\begin{aligned} E[X^2] - E[X] &= \mu^2, \\ E[X^2] &= \mu^2 + \mu, \\ V[X] = E[X^2] - (E[X])^2 &= \mu. \end{aligned}$$

Deriving the mean and variance for the binomial distribution follows the procedure used in (4.12) and (4.13), and it is left as an exercise to show that the expectation and variance for a binomial is given by (4.14):

4

$$\begin{aligned} E[X] &= Np, \\ V[X] &= Np(1-p). \end{aligned} \tag{4.14}$$

The manipulations used to arrive at (4.14) are relatively involved, compared with an alternative method based on generating functions—this is discussed in 4.6.

4.4 Folding of probability density functions

Here we compute pulse height distributions, at the output of a PMT, for two different forms of the SER. The first is based on an exponential pulse height distribution and is representative of a PMT with a poor SER; the second concerns an SER of an actual PMT showing a valley and a well-defined peak in the distribution—this is representative of a good quality PMT. The SER for an exponential distribution has the following form:

$$p_0(h) = ae^{-ah}. \tag{4.15}$$

Equation (4.15) is normalized since $\int_0^\infty p_0(h)dh = 1$. The two SERs are shown in Fig. 4.2.

The mean, m , and variance, var , of (4.15) are, respectively,

$$m = \frac{\int_0^\infty ae^{-ah}h dh}{\int_0^\infty ae^{-ah} dh} = \frac{1}{a}, \tag{4.16}$$

and

$$\text{var} = \frac{\int_0^\infty ae^{-ah}(h-m)^2 dh}{\int_0^\infty ae^{-ah} dh} = \frac{1}{a^2}, \tag{4.17}$$

giving $\sigma = 1/a$, and $\sigma/m = 1$.

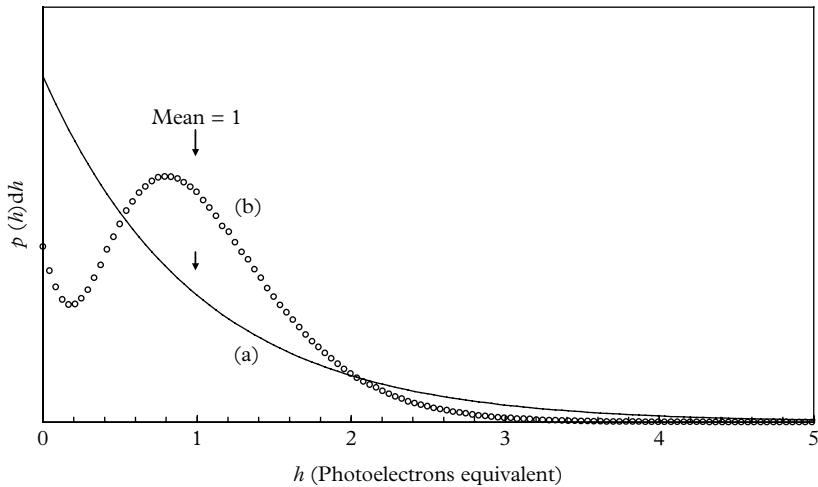


Fig. 4.2. An SER is the pulse height distribution measured by exciting a PMT with single photons. The solid curve, (4.15), is an exponential distribution, and the circles refer to a measured SER. Distribution (b) has been scaled so that the mean appears at 1 photoelectron equivalent.

For the distribution labelled (b) in Fig. 4.2, the mean and the standard deviation are calculated directly from experimental coordinates, to arrive at $\sigma/m = 0.61$. The information sought is the shape of the spectrum corresponding to an input of exactly n pe, where (4.15) is the individual photoelectron response. In this case, the problem is entirely analytical, whereas with the experimental SER assumed in (b), recourse to Monte Carlo simulation is necessary. Using the folding method we can generate pulse height distributions for any number of simultaneous photoelectrons originating from a defined distribution. Consider the distribution for two photoelectrons escaping the photocathode simultaneously. If the combined pulse height is H and if one of the photoelectrons produces a pulse of height h , then the other photoelectron will produce a pulse of height $(H-h)$. The combined distribution is thus

$$P(m) = \int_0^m p(\nu)q(m-\nu)d\nu. \quad (4.18)$$

Since we are selecting two photoelectrons from the same input distribution, the functions p and q are identical and may be replaced by p_0 , with the subscript identifying the spectrum to be folded. The subscript 1 indicates the first folding, corresponding to two electrons, so that (4.18) becomes

$$p_1(H) = \int_0^H p_0(h)p_0(H-h)dh.$$

Assuming an SER of the form $p_0(h) = ae^{-ah}$, the probability distribution for an input of 2 pe is

$$\begin{aligned} p_1(H) &= a^2 \int_0^H e^{-ah} \times e^{-a(H-h)} dh \\ &= a^2 e^{-aH} H. \end{aligned}$$

4

Note that H is a constant since the integration variable is h and hence e^{-aH} may be taken outside this integral. Furthermore, H is a dummy variable (we can call it anything we like) and the convention is to replace it by h to give $p_1(h) = a^2 h e^{-ah}$, which is the distribution for 2 pe. The distribution for 3 pe follows by folding the distribution for 2 pe with that for 1 pe:

$$\begin{aligned} p_2(H) &= \int_0^H a^2 h e^{-ah} a e^{-a(H-h)} dh \\ &= \frac{a^3 e^{-aH} H^2}{2!}, \\ p_2(h) &= \left(\frac{a^3 e^{-ah} h^2}{2!} \right). \end{aligned} \quad (4.19)$$

The n -fold distribution, for $(n+1)$ pe, is

$$p_n(h) = \frac{a^{n+1}}{n!} e^{-ah} h^n, \quad (4.20)$$

$$m = \frac{\int_0^\infty p_n(h) h dh}{\int_0^\infty p_n(h) dh} = \frac{n+1}{a}. \quad (4.21)$$

The variance is

$$\text{var}_n = \frac{\int_0^\infty p_n(h) (h - m_n)^2 dh}{\int_0^\infty p_n(h) dh} = \frac{n+1}{a^2} \quad (4.22)$$

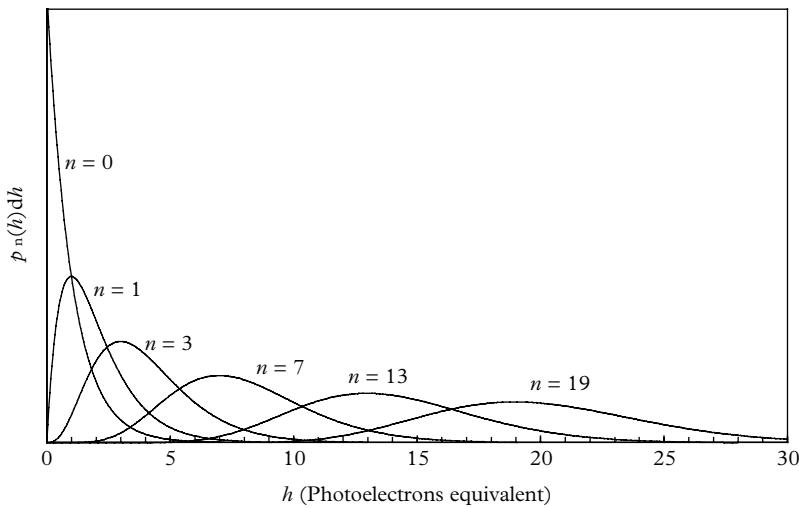


Fig. 4.3. Multiple-photoelectron distributions derived from an exponential SER, labelled $n = 0$. The 2 pe distribution is given by $n = 1$ and $m = 2$; in general, the curve for $(n + 1)$ pe will follow the distribution labelled n . All distributions with $n \geq 1$ show a peak, even though the fundamental curve, $n = 0$, does not. This illustrates the central limit theorem, with the distributions becoming more symmetrical (normal-like) with increasing n .

and

$$\frac{\sigma_n}{m_n} = \frac{1}{(n+1)^{1/2}} .$$

Note in (4.21), for $n = 0$, $m_0 = 1/a$ and, in (4.22), $\text{var}_0 = 1/a^2$. Equation (4.20) is plotted in Fig. 4.3 for the indicated n values.

4.5 Moments

4.5.1 Moments about the origin

By analogy with statics, $\langle x \rangle$, as previously defined, is also known as the first moment about the origin of the distribution, $p(x)$, invariably referred to as the mean, μ . Generalizing,

$$\langle x^k \rangle = \sum_x x^k p(x) = M_k \quad (4.23)$$

is the k^{th} moment of a distribution, noting that the qualification ‘about the origin’ is often omitted, but understood to apply. It is instructive to evaluate (4.23) for a

Poisson distribution, starting with the first moment, which has already been stated in (4.8) and (4.13) as

$$\langle x \rangle = \sum_x x p(x) = \mu = M_1.$$

The second moment is

$$\langle x^2 \rangle = \sum_x x^2 p(x) = \sum_x x^2 \frac{e^{-\mu} \mu^x}{x!} = \mu + \mu^2 = M_2.$$

The expression for the third moment is not so easily derived by these methods but a recurrence formula is given by Jánossy (1965); from this, $\langle x^3 \rangle$ follows as

$$\langle x^3 \rangle = \mu + 3\mu^2 + \mu^3 = M_3.$$

The first four moments for a Poisson distribution are

$$\begin{aligned} M_1 &= \langle x \rangle = \mu, \\ M_2 &= \langle x^2 \rangle = \mu + \mu^2, \\ M_3 &= \langle x^3 \rangle = \mu + 3\mu^2 + \mu^3, \\ M_4 &= \langle x^4 \rangle = \mu + 7\mu^2 + 6\mu^3 + \mu^4. \end{aligned}$$

4.5.2 Central moments

Moments taken about the mean are called ‘central moments’, denoted by m_k . The first central moment is

$$m_1 = \sum_x (x - \mu) p(x) = 0.$$

The second central moment, m_2 , recognized as the variance, follows from (4.6) and (4.7) as

$$m_2 = \sum_x (x - \mu)^2 p(x) = \langle x^2 \rangle - \mu^2 = \text{var}(X).$$

The third central moment is

$$\begin{aligned} m_3 &= \sum_x (x - \mu)^3 p(x) \\ &= \sum_x (x^3 - 3\mu x^2 + 3\mu^2 x - \mu^3) p(x) \\ &= M_3 - 3\mu M_2 + 3\mu^2 M_1 - \mu^3 \end{aligned}$$

or, since $\mu \equiv M_1$,

$$m_3 = M_3 - 3M_1 M_2 + 2M_1^3.$$

The entity, γ , known as the skew, is defined as $\gamma = m_3/\sigma^3$, where σ^2 is the variance of the distribution. The skew is zero if the data are symmetrically distributed about the mean. If a tail extends to the right, γ is positive and the data have a positive skew. If the tail is to the left, the skew is negative. Skew is seldom used in applications involving

PMTs, although negative skew in a peaked pulse height distribution is usually indicative of non-linearity. The first five central moments for a Poisson distribution are

$$\begin{aligned} m_0 &= \sum_x (x - \mu)^0 p(x) = \sum_x p(x) = 1, \\ m_1 &= \sum_x (x - \mu)^1 p(x) = \sum_x x p(x) - \mu \sum_x p(x) = 0, \\ m_2 &= \sum_x (x - \mu)^2 p(x) = \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) = \mu, \\ m_3 &= \sum_x (x - \mu)^3 p(x) = \mu, \\ m_4 &= \sum_x (x - \mu)^4 p(x) = \mu + 3\mu^2. \end{aligned}$$

4

A recurrence formula is also given by Jánossy, for the central moments.

4.5.3 Factorial moments

A third set of computations for use with statistical data, referred to as ‘factorial moments’, is defined as

$$\langle x^{[k]} \rangle = \sum_{x=0}^{\infty} x^{[k]} p(x),$$

where $x^{[k]}$ denotes the factorial $\langle x(x-1)(x-2)\cdots(x-k+1) \rangle$, $k = 1, 2, 3, \dots$. This entity is defined as

$$\begin{aligned} \frac{\langle x^{[k]} \rangle}{\langle x \rangle^k} &= \frac{\langle x(x-1)(x-2)\cdots(x-k+1) \rangle}{\langle x \rangle^k} p(x); \\ \frac{\langle x^{[1]} \rangle}{\langle x \rangle} &= \sum_x \frac{x p(x)}{\langle x \rangle} \\ &= \frac{M_1}{M_1} = 1, \\ \frac{\langle x^{[2]} \rangle}{\langle x \rangle^2} &= \sum_x \frac{x(x-1)}{\langle x \rangle^2} p(x) = \sum_x \frac{(x^2 - x)}{\langle x \rangle^2} p(x) \\ &= \frac{(M_2 - M_1)}{M_1^2}, \\ \frac{\langle x^{[3]} \rangle}{\langle x \rangle^3} &= \sum_x x \frac{(x-1)(x-2)}{\langle x \rangle^3} p(x) = \sum_x \frac{(x^3 - 3x^2 + 2x)}{\langle x \rangle^3} p(x) \\ &= \frac{(M_3 - 3M_2 + 2M_1)}{M_1^3}. \end{aligned}$$

For a Poisson distribution, the first normalized factorial moment is obviously unity. The second is easily uncovered,

$$\frac{\langle x^{[2]} \rangle}{\langle x \rangle^2} = \frac{\langle x(x-1) \rangle}{\langle x \rangle^2} = \frac{\langle x^2 \rangle - \langle x \rangle}{\langle x \rangle^2} = \frac{\mu + \mu^2 - \mu}{\mu^2} = 1,$$

as is the third:

$$\begin{aligned}\frac{\langle x^{[3]} \rangle}{\langle x \rangle^3} &= \frac{\langle x(x-1)(x-2) \rangle}{\langle x \rangle^3} = \frac{\langle x^3 \rangle - 3\langle x^2 \rangle + 2\langle x \rangle}{\langle x \rangle^3} \\ &= \frac{\mu + 3\mu^2 + \mu^3 - 3(\mu + \mu^2) + 2\mu}{\mu^3} = 1.\end{aligned}$$

All normalized factorial moments are unity for a Poisson distribution; these formulations prove to be particularly useful for testing whether a PMT is statistically well behaved under photon-counting conditions (see 4.5.4). The fourth and higher-order moments follow by adopting the same procedure above, but the manipulations become unwieldy.

4.5.4 Exercise in photon counting

The statistical considerations so far covered are applied to a set of repeated count intervals derived from a low level single-photon source. The exercise presented is based on a count rate of nominally 1000 per second with a negligible background contribution. A total of 10,000 repeated samples, each taken over a 10 ms time interval, form the probability density function (pdf) recorded in Table 4.1 and

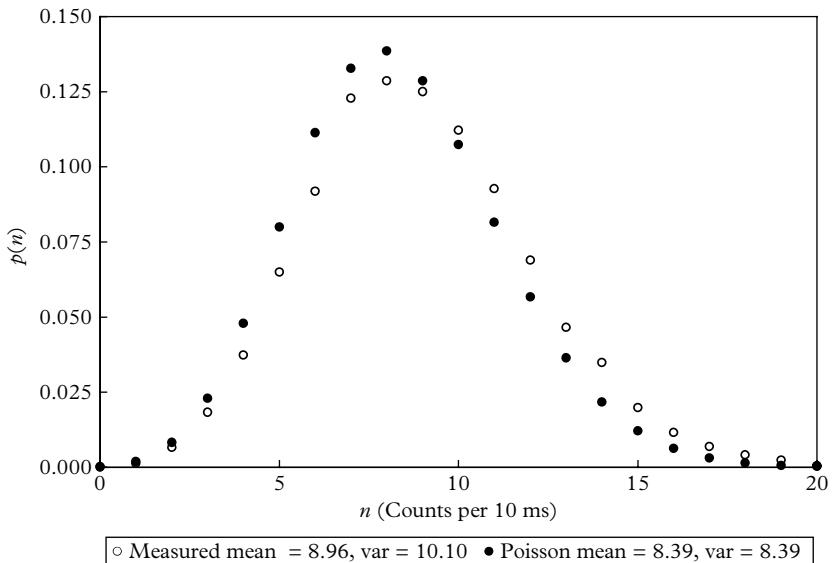


Fig. 4.4. The pdf for repeated exposure to a random but steady source of counts, extracted from Table 4.1 (○). The mean count rate per 10 ms is 8.96, with a variance of 10.10. An ideal PMT, free from afterpulses, is represented by the Poisson distribution, with mean and variance in the region of 8.39 (●).

Table 4.1 A set of 10,000 repeated signal counts of a random source illustrates the statistical parameters given in the text. The measured mean count rate is 8.959 per 10 ms, corrected to 8.39 after subtraction of afterpulses. The formulation for the afterpulse rate is based on (4.99). Factorial moments are unity for a Poisson distribution, as shown in 4.5.3. Any deviation from unity is a measure of the statistical quality of a PMT and serves as a means for grading performance. Entries in the last column refer to a mean of 8.39.

x	$f(x)$	$p(x)$	$x p(x)$	$(x-\mu)^2 p(x)$	M_1	M_2	M_3	M_4	Poisson
0	1	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000
1	13	0.001	0.001	0.082	0.001	0.001	0.001	0.001	0.002
2	66	0.007	0.013	0.320	0.013	0.026	0.053	0.106	0.008
3	183	0.018	0.055	0.650	0.055	0.165	0.494	1.482	0.023
4	373	0.037	0.149	0.917	0.149	0.597	2.387	9.549	0.048
5	650	0.065	0.325	1.019	0.325	1.625	8.125	40.625	0.080
6	919	0.092	0.551	0.805	0.551	3.308	19.850	119.102	0.111
7	1228	0.123	0.860	0.471	0.860	6.017	42.120	294.843	0.133
8	1286	0.129	1.029	0.118	1.029	8.230	65.843	526.746	0.139
9	1250	0.125	1.125	0.000	1.125	10.125	91.125	820.125	0.129
10	1121	0.112	1.121	0.121	1.121	11.210	112.100	1121.000	0.107
11	927	0.093	1.020	0.386	1.020	11.217	123.384	1357.221	0.081
12	689	0.069	0.827	0.637	0.827	9.922	119.059	1428.710	0.057
13	465	0.047	0.605	0.759	0.605	7.859	102.161	1328.087	0.036
14	349	0.035	0.489	0.887	0.489	6.840	95.766	1340.718	0.022
15	199	0.020	0.299	0.726	0.299	4.478	67.163	1007.438	0.012
16	116	0.012	0.186	0.575	0.186	2.970	47.514	760.218	0.006
17	69	0.007	0.117	0.446	0.117	1.994	33.900	576.295	0.003
18	41	0.004	0.074	0.335	0.074	1.328	23.911	430.402	0.001
19	24	0.002	0.046	0.242	0.046	0.866	16.462	312.770	0.001
20	5	0.001	0.010	0.061	0.010	0.200	4.000	80.000	0.000
21	14	0.001	0.029	0.203	0.029	0.617	12.965	272.273	0.000
22	7	0.001	0.015	0.119	0.015	0.339	7.454	163.979	0.000
23	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
26	1	0.000	0.003	0.029	0.003	0.068	1.758	45.698	0.000
27	1	0.000	0.003	0.033	0.003	0.073	1.968	53.144	0.000

(Continued)

Table 4.1 Continued

<i>x</i>	<i>f(x)</i>	<i>p(x)</i>	<i>x p(x)</i>	$(x-\mu)^2 p(x)$	<i>M</i> ₁	<i>M</i> ₂	<i>M</i> ₃	<i>M</i> ₄	Poisson
28	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
30	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
31	2	0.000	0.006	0.097	0.006	0.192	5.958	184.704	0.000
32	1	0.000	0.003	0.053	0.003	0.102	3.277	104.858	0.000
$\sum f(x)$		$\sum p(x)$	μ	<i>var</i>	<i>M</i> ₁	<i>M</i> ₂	<i>M</i> ₃	<i>M</i> ₄	Poisson
10000		1.000	8.959	10.101	8.959	90.370	1008.797	12380.093	1.000

$$M_1 = 8.96$$

$$M_2 = 90.37$$

$$M_3 = 1008.80$$

$$M_4 = 12380.09$$

$$\begin{aligned} \alpha^{[1]}/\alpha^1 &= M_1/M_1 &= 1.000 \\ \alpha^{[2]}/\alpha^2 &= (M_2 - M_1)/M_1^2 &= 1.014 \\ \alpha^{[3]}/\alpha^3 &= (M_3 - 3M_2 + 2M_1)/M_1^3 &= 1.051 \\ \alpha^{[4]}/\alpha^4 &= (M_4 - 6M_3 + 11M_2 - 6M_1)/M_1^4 &= 1.128 \\ \text{Afterpulse rate} &= (\text{var}/\mu - 1)/(3 - \text{var}/\mu) &= 6.40\% \\ \text{True signal rate/10 ms} &= 8.959(1 - 0.064) &= 8.39 \end{aligned}$$

presented graphically in Fig. 4.4. Such a data set takes only 100 s to acquire. Drift in PMT gain or drift in the input light level has minimal effect over such a short accumulation time. The method therefore offers a means to examine the statistical behaviour of a PMT rather than its long-term instability. Of course, if the latter is also of interest, then the exercise may be repeated with longer sample times. The mean, variance, and the first four factorial moments are listed, using the formulations already given, with the presumption that the experimental distribution will be Poisson. However, it is obvious that the variance of the experimental data exceeds the mean, primarily because of correlations in the data, principally afterpulses, which cause a shift in $f(x)$ and $p(x)$ to the right in the pdf. It is shown in 4.5.3 that the normalized factorial moments are unity for a Poisson distribution but, as is the case for the calculated variance, the moments derived from the data all exceed unity. It is the user's prerogative to set the level of acceptable performance. This is achieved by establishing the optimal sensitivity through the choice of PMT gain, counting threshold, and preamplifier gain (if included). Factorial moments also serve as a means for grading PMT performance.

4.6 Probability generating functions

Generating functions, developed by Laplace, provide an elegant means for calculating probabilities and variances of complex statistical distributions. The statistics of a system where Process A is followed by Process B is complicated to quantify using conventional statistical methods of the previous sections. As an example of linked processes in PMTs, A may represent the statistical process of converting incoming photons to photoelectrons followed by B, the multiplication produced by the dynodes. If we have a multiplier of n dynodes, then Process AB involves $(1 + n)$ sequential and independent statistical processes. It will be obvious that the use of generating functions offers an elegant and powerful means for handling such statistical processes.

Consider any probability distribution $p(x)$, where x may be finite or infinite. In the case of a die, x ranges from 1 to 6 but, for a Poisson distribution, x extends to infinity: that is, all outcomes are possible. The generating function $G(u)$ for the distribution $p(x)$ is defined as

$$G(u) = p(0)u^0 + p(1)u^1 + p(2)u^2 + \dots + p(x)u^x + \dots, \quad (4.24)$$

with $G(1) = 1$.

It can be noted from (4.24) that $G(u)$ embodies all the probabilities of a distribution—hence the terminology ‘probability generating function’; $G(u)$ can also be interpreted as the expected value u^X . We regard the $p(x)$ functions as coefficients of the ancillary variable u , with $G(u)$ a proper function of u ; $G'(u)$ is the first derivative of $G(u)$ with respect to u only, and some texts use the partial differentiation symbol to emphasize this point. Differentiating (4.24) leads to

$$\begin{aligned} G'(u) &= p(1)u^0 + 2p(2)u^1 + 3p(3)u^2 + \dots \\ &= \sum_{x=0}^{\infty} x p(x) u^{x-1}, \\ G'(1) &= \sum_{x=0}^{\infty} x p(x) = E[X]. \end{aligned} \quad (4.25)$$

Thus, $G'(1)$ gives the mean or expectation, $E[X]$. Differentiating again yields

$$G''(u) = \sum_{x=0}^{\infty} x(x-1)p(x)u^{x-2},$$

from which follows

$$G''(1) = \sum_{x=0}^{\infty} x(x-1)p(x) = E[X(X-1)]. \quad (4.26)$$

An expression for the variance can be obtained from a combination of the derivatives of $G(u)$. Consider the expression

$$\begin{aligned} G''(1) + G'(1) - (G'(1))^2 &= E[X^2] - E[X] + E[X] - (E[X])^2 \\ &= E[X^2] - (E[X])^2, \end{aligned}$$

which is the variance, and therefore

$$\text{var}(X) = G''(1) + G'(1) - (G'(1))^2. \quad (4.27)$$

The relative ease with which the generating function, as presented in (4.24), provides the mean and variance of any probability distribution is demonstrated for the Poisson distribution; thus,

$$\begin{aligned} G(u) &= \sum_x e^{-\mu} \frac{\mu^x}{x!} u^x = \sum_x e^{-\mu} \frac{(\mu u)^x}{x!} = e^{-\mu} \sum_x \frac{(\mu u)^x}{x!} = e^{-\mu} e^{\mu u} = \exp(\mu(u - 1)) \\ G'(u) &= \exp(\mu(u - 1))\mu, \\ G'(1) &= \mu, \\ G''(u) &= \exp(\mu(u - 1))\mu^2, \\ G''(1) &= \mu^2. \end{aligned} \quad (4.28)$$

The expectation value, the mean, is given by $G'(1) = \mu$, and the variance follows from (4.27):

$$\text{var}(X) = \mu^2 + \mu - \mu^2 = \mu.$$

The variance of a binomial distribution follows equally easily by this method:

$$\begin{aligned} p(x) &= {}^N C_x p^x (1-p)^{N-x} \\ G(u) &= \sum_x u^x p(x) = \sum_{x=0}^N {}^N C_x (up)^x (1-p)^{N-x}. \end{aligned}$$

We recognize that the right-hand side is the binomial expansion of $[up + (1-p)]^N$, so

$$\begin{aligned} G(u) &= [1 + (u - 1)p]^N, \\ G'(u) &= N[1 + (u - 1)p]^{N-1}p, \\ G'(1) &= Np, \\ G''(u) &= N(N-1)[1 + (u - 1)p]^{N-2}, \\ G''(1) &= N(N-1)p^2. \end{aligned} \quad (4.29)$$

The variance follows from (4.27) as

$$\text{var}(X) = N(N-1)p^2 + Np - N^2p^2 = Np(1-p). \quad (4.30)$$

The formulations (4.25) and (4.27) are of exceptional practical utility for solving complicated combinatorial manipulations. This will be apparent when analysing the combined action of two or more independent statistical processes, discussed in 4.7.

4.7 Sum of chance integers

We now consider statistical processes that are linked and the way to handle them using generating functions. First, consider two independent statistical events the outcomes of which are added—an everyday example of this is tossing two dice. There is an analogous situation with PMTs when measuring signal in the presence of background. In photon counting, for example, a source signal, S , is measured concurrently with an undesired background, B . The signal and background counts are independent and we measure the sum of the two processes in any counting exercise. Measurement of B alone can be made for subtraction, but we cannot measure S directly, unless $B = 0$. Consider two independent probability distributions, $p(k)$ and $q(j)$, which may or may not be of the same type, and write S_{p+q} as the sum of a dual outcome:

$$S_{p+q} = p(k) + q(j).$$

The probability $P(m)$ of recording m counts within a time interval t is given by considering all possible combinations of $p(k)$ and $q(j)$ that generate m counts. The sum of the two components can only be zero if both are zero; thus,

$$P(0) = p(0)q(0).$$

For $m = 1$ we have

$$P(1) = p(0)q(1) + p(1)q(0).$$

In general we can obtain a sum equal to m from all product pairs:

$$p(0)q(m), \quad p(1)q(m-1), \quad \dots, \quad p(m)q(0),$$

leading to

$$P(m) = \sum_{\nu=0}^m p(\nu) \times q(m-\nu). \quad (4.31)$$

Combining independent variables in the manner of (4.31) is commonplace, in both statistical and non-statistical applications. This is called folding or convolution of two variables and is represented symbolically as

$$P(m) = p(k) \otimes q(j).$$

We seek to express (4.31) in terms of generating functions, in anticipation that this will aid in the manipulation of combined statistical processes. Two generating functions, one for $p(k)$ and one for $q(j)$, are defined respectively as

$$G_p(u) = \sum_{\nu=0}^{\infty} u^{\nu} p(\nu).$$

and

$$G_q(u) = \sum_{\nu=0}^{\infty} u^{\nu} q(\nu).$$

The generating function for the compound distribution $P(m)$ is denoted by $G_P(u)$ and, using (4.31),

$$G_P(u) = \sum_{\nu=0}^{\infty} u^{\nu} P(\nu) = \sum_{m=0}^{\infty} u^m \sum_{\nu=0}^m p(\nu) q(m-\nu). \quad (4.32)$$

Inserting $m = \nu + s$ in (4.32) yields

$$\begin{aligned} G_P(u) &= \sum_{\nu,s=0}^{\infty} u^{\nu+s} \sum_{\nu=0}^m p(\nu) q(m) \\ &= \sum_{\nu,s=0}^{\infty} u^{\nu} p(\nu) \times u^s q(s) \\ &= G_p(u) \times G_q(u). \end{aligned} \quad (4.33)$$

The combined statistics of two Poisson processes, for example, with means of μ_1 and μ_2 have a generating function given by (4.33):

$$\begin{aligned} G_P(u) &= \exp(\mu_1(u - 1)) \times \exp(\mu_2(u - 1)) \\ &= \exp((\mu_1 + \mu_2)(u - 1)), \end{aligned} \quad (4.34)$$

which is the generating function of a Poisson distribution with a mean of $(\mu_1 + \mu_2)$.

Differentiating (4.34), we obtain

$$G_P'(u) = G_P(u)(\mu_1 + \mu_2),$$

from which

$$G_P'(1) = \mu_1 + \mu_2,$$

$$G_P''(u) = G_P(u)(\mu_1 + \mu_2)^2,$$

from which

$$G_P''(1) = (\mu_1 + \mu_2)^2,$$

so

$$\text{var} = G_P''(1) + G_P'(1) - (G_P'(1))^2 = \mu_1 + \mu_2,$$

which we knew anyway. The standard deviation, σ , is the square root of the variance, by definition, and from this we have the important result: variances add linearly, while standard deviations add in quadrature for Poisson and normal distributions:

$$\text{var} = \mu_1 + \mu_2,$$

$$(\sigma_{1+2})^2 = \sigma_1^2 + \sigma_2^2.$$

This folding process does not alter the functional dependence of the output. Equation (4.33) is simply the product of the two generating functions, but it is important to stress that the folding process only has meaning when the two variables are independent. Equation (4.33) is quite general and refers to any two statistical distributions whether Poisson, binomial, or whatever. Folding can be extended to multiple random variables; thus,

$$G_{1,2,\dots,n}(u) = G_1(u) \times G_2(u) \times \dots \times G_n(u).$$

If all n distributions have the same statistical form (all Poisson for example), we can express $G_{1,2,\dots,n}(u)$ as follows:

$$G_{1,2,\dots,n}(u) = \left(G_1(u) \right)^n. \quad (4.35)$$

4.8 Moment generating functions

Certain statistical manipulations are more amenable to a different form of generating function. Consider the expansion, $G(\nu)$, in the manner of (4.24) with u replaced by e^ν :

$$G(\nu) = p(0)e^0 + p(1)e^\nu + p(2)e^{2\nu} + \dots + p(x)e^{x\nu} + \dots \quad (4.36)$$

Here $G(\nu)$ is the expected value of $e^{x\nu}$ and, as before, $G(\nu) = \langle e^{x\nu} \rangle$; it is sometimes called the moment generating function, because that is what it does. First, we derive $G(\nu)$ for a Poisson distribution by substituting the appropriate terms into (4.36):

$$\begin{aligned} G(\nu) &= e^{-\mu} \left(1 + \mu e^\nu + \frac{(\mu e^\nu)^2}{2!} + \frac{(\mu e^\nu)^3}{3!} + \dots \right) \\ &= e^{-\mu} (\exp(\mu e^\nu)) \\ &= \exp(\mu(e^\nu - 1)). \end{aligned} \quad (4.37)$$

Note that (4.37) and (4.28) have the same form. The derivatives of (4.36) with respect to ν are

$$G'(\nu) = p(1)e^\nu + 2p(2)e^{2\nu} + \dots + xp(x)e^{x\nu} = \sum_{x=0}^{\infty} xp(x)e^{x\nu},$$

$$G''(\nu) = 1^2 p(1)e^\nu + 2^2 p(2)e^{2\nu} + \dots + x^2 p(x)e^{x\nu} = \sum_x x^2 p(x)e^{x\nu},$$

$$G^{(k)}(\nu) = \sum_x x^k p(x)e^{x\nu},$$

where (k) signifies the k^{th} differential of the generating function, with k an index:

$$G^{(k)}(0) = \sum_x x^k p(x) = M_k.$$

Here M_k is the k^{th} moment, as discussed in 4.8, with the useful result that the derivatives of the moment generating function $G(\nu)$ are equal to the moments of the distribution for $\nu = 0$. The first two derivatives are

$$G'(0) = \sum_{x=0}^{\infty} x p(x) = E[X]$$

and

$$G''(0) = E[X^2],$$

leading to

$$\text{mean} = G'(0),$$

$$\text{var}(X) = G''(0) - (G'(0))^2. \quad (4.38)$$

$\text{var}(X)$ can be determined for a Poisson distribution by using $G(\nu)$, as follows.

Starting with the generating function for a Poisson distribution, (4.37), and differentiating, we have

$$\begin{aligned} G(\nu) &= \exp(\mu(e^\nu - 1)), & M_0 &= G(0) = 1; \\ G'(\nu) &= \exp(\mu(e^\nu - 1))\mu e^\nu = G(\nu) \mu e^\nu, & M_1 &= G'(0) = \mu; \\ G''(\nu) &= G'(\nu)\mu e^\nu + G(\nu)\mu e^\nu, & M_2 &= G''(0) = \mu^2 + \mu; \\ G'''(\nu) &= G''(\nu)\mu e^\nu + G'(\nu)\mu e^\nu + G'(\nu)\mu e^\nu + G(\nu)\mu e^\nu, & M_3 &= G'''(0) = (\mu^2 + \mu)\mu + 2\mu^2 + \mu. \end{aligned}$$

Higher-order moments are derived by continuing this procedure or, more conveniently, by using the recurrence formula

$$\begin{aligned} G^{(k)}(0) &= M_L = \mu \sum_{k=0}^{L-1} \frac{(L-1)!}{k!(L-1-k)!} M_k; \\ \therefore \text{var}(X) &= G''(0) - (G'(0))^2 = M_2 - M_1^2 = \mu. \end{aligned} \quad (4.39)$$

The reader can verify, by following steps (4.36) to (4.39), that generating functions defined in terms of the complex variable $i\nu$ lead to the relationships

Table 4.2 Generating and statistical parameters for a family of exponential distributions.

Variable	Generating function	$E[X]$	$\text{var}(X)$
e^ν	$\exp(\mu \times (e^\nu - 1))$	$G'(0)$	$G''(0) - (G'(0))^2$
$e^{-\nu}$	$\exp(\mu \times (e^{-\nu} - 1))$	$-G'(0)$	$G''(0) - (G'(0))^2$
$e^{i\nu}$	$\exp(\mu \times (e^{i\nu} - 1))$	$-iG'(0)$	$-G''(0) + (G'(0))^2$
$e^{-i\nu}$	$\exp(\mu \times (e^{-i\nu} - 1))$	$iG'(0)$	$-G''(0) + (G'(0))^2$

shown in Table 4.2, all of which generate the same mean and variance, whatever the particular distribution. This is readily checked for the Poisson and binomial distributions, or any others, by using the formulations in the table. The results are independent of the form adopted for the variable in the generating function, which emphasizes that they serve purely as a means to an end. Obviously, one should choose the form that best suites the problem the exponential type is particularly suited to continuous probability distributions, for example.

Analogously to (4.10) and (4.11), we have

$$\langle e^{-\nu x} \rangle = \int e^{-\nu x} p(x) dx = f(\nu) = L(p(x))$$

and

$$\langle e^{-i\nu x} \rangle = \int e^{-i\nu x} p(x) dx = f(\nu) = F(p(x)),$$

where $L(p(x))$ and $F(p(x))$ have the familiar forms of the Laplace and the Fourier transforms, respectively.

4.9 Folding with probability generating functions

Following (4.35), the generating function for a distribution $p_0(h)$ folded with itself n times is given by

$$G_n(\nu) = (G_0(\nu))^n,$$

where $G_0(\nu)$, for example, is the generating function of $p_0(h) = ae^{-ah}$, given by

$$\begin{aligned} G_0(\nu) &= \int_0^{\infty} e^{\nu h} a e^{-ah} dh = \frac{a}{(a-\nu)}, \\ G_n(\nu) &= \frac{a^{n+1}}{(a-\nu)^{n+1}}, \\ G'_n(\nu) &= \frac{(n+1)}{a} \frac{a^{n+2}}{(a-\nu)^{n+2}}. \end{aligned}$$

4

The mean on the n^{th} folded distribution is therefore

$$G'_n(0) = \frac{n+1}{a}, \quad (4.40)$$

giving

$$\begin{aligned} G''_n(0) &= \frac{(n+1)(n+2)}{a^2}, \\ \text{var}_n &= G''_n(0) - (G'_n(0))^2 = \frac{n+1}{a^2}. \end{aligned} \quad (4.41)$$

Equations (4.40) and (4.41) are identical to (4.21) and (4.22), as they must be. Note, however, that the distributions $p_n(h)$ are not revealed in this treatment, as they are in (4.20).

The folding process is repeated for the peaked SER in Fig. 4.2. The distributions in Fig. 4.5 are created by a Monte Carlo selection method. Profiles are

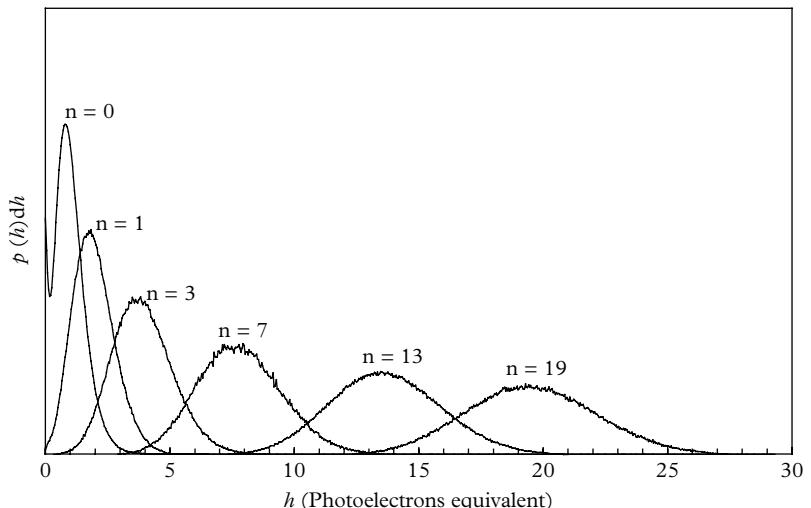


Fig. 4.5. Multiple-photoelectron distributions derived from a peaked SER. The noise on the curves arises from the statistics of the adopted Monte Carlo process.

generated, for the case $n=3$, for example, by selecting four pulse heights randomly from the SER followed by summing; thus,

$$h = h_0 + h_1 + h_2 + h_3.$$

This process is repeated many times, for any n , until the response curves assume a recognizable shape. The improvement in resolution over that for an exponential SER is marked. For example, the full widths of the two distributions for $n=19$ differ, approximately, by a factor of 2: this underlines the importance of a narrow single-electron distribution, which is the point of the exercise. The purpose of this section is to illustrate folding and the importance of a well-resolved SER, bearing in mind that the depictions in Figs 4.3 and 4.5 are artificial since no allowance has been made for fluctuations at the photocathode. For example, it is possible to produce a photoelectron source that gives pulses of 19 pe on average, as assumed for n , but never exactly 19 pe every time.

The area under each curve in Figs 4.3 and 4.5 is unity; that is, each curve has the same number of events. We can easily see that the peak heights in Fig. 4.5 exceed the peak heights in Fig. 4.3, an obvious sign that the multi-photoelectron resolution is better for the distributions in Fig. 4.5. This brings out the importance of a low-noise SER with regard to resolution—the better the SER, the better is the multi-photoelectron resolution.

4.10 Cascaded processes

Cascaded processes are those in which the outcome of the first statistical event serves as the input to a second statistical process. Event A occurs first, giving a score a ; then Event B occurs a times; and then the scores of these trials are added to give a total score. Cascade theory developed here is universal in the sense that the formulations for the mean and variance do not depend on the type of distribution, requiring only the mean and the variance, as will become obvious. An example of a two-stage cascade is the phenomenon of light emission followed by photodetection. The theory can be extended to any number of cascades to cover the behaviour of a k -stage multiplier, for example.

4.10.1 Generalized two-stage cascading

The generating function for a two-stage cascade has the form

$$\begin{aligned} G_{AB}(u) &= G_A[G_B(u)], \\ G_{AB}'(u) &= G_A'[G_B(u)] \times G_B'(u). \end{aligned} \tag{4.42}$$

Recall from (4.24) and (4.25) that the generating functions for A and B have the following properties:

$$\begin{aligned} G_A(1) &= 1, \\ G'_A(1) &= m_A, \\ G_B(1) &= 1, \\ G'_B(1) &= m_B. \end{aligned}$$

4

From (4.7) and (4.9), for any generating function,

$$\begin{aligned} G''(1) &= \langle u^2 \rangle - \langle w \rangle, \\ \text{var} &= \langle u^2 \rangle - \langle w \rangle^2; \\ \therefore G''_A(1) &= \text{var}_A + m_A^2 - m_A, \\ G''_B(1) &= \text{var}_B + m_B^2 - m_B \\ G'_{AB}(1) &= G'_A(1)G'_B(1) = m_A \times m_B. \end{aligned} \tag{4.43}$$

Differentiating (4.42) and setting $u = 1$,

$$\begin{aligned} G_{AB}''(u) &= G''_A[G_B(u)] \times (G'_B(u))^2 + G'_A(u) \times G''_B(u) \\ G_{AB}''(1) &= G''_A(1) \times (G'_B(1))^2 + G'_A(1) \times G''_B(1) \\ &= (\text{var}_A + m_A^2 - m_A) \times m_B^2 + m_A \times (\text{var}_B + m_B^2 - m_B) \\ &= \text{var}_A \times m_B^2 + \text{var}_B \times m_A + (m_A \times m_B)^2 - m_A \times m_B; \\ \text{var}_{AB} &= G_{AB}''(1) + G'_{AB}(1) - (G'_{AB}(1))^2, \text{ from (4.27);} \\ \therefore \text{var}_{AB} &= \text{var}_A \times m_B^2 + \text{var}_B \times m_A \end{aligned} \tag{4.44}$$

4.10.2 Light detection statistics

We now apply (4.44) to a problem for which the answer is already familiar—that of photodetection. As before, we consider a light source emitting on average m_p photons per second, and Poisson distributed with a variance of m_p . Therefore, in (4.44) var_A is taken as m_p and, similarly, m_A is replaced by m_p . The conversion to photoelectrons at the photocathode is governed by the binomial distribution, for which var_B is $\eta(1 - \eta)$, and m_B becomes η . Substituting into (4.44) gives the variance for photoelectron emission, var_{pe} , while (4.43) predicts the mean photo-electron rate, $m_p\eta$. The variance for the combined process is

$$\begin{aligned} \text{var}_{pe} &= m_p\eta^2 + \eta(1 - \eta)m_p \\ &= m_p\eta. \end{aligned} \tag{4.45}$$

The mean number of photoelectrons is given by 4.43 as $m_A \times m_B = m_p \times \eta$. The variance and the mean of the combined process are identical: hence the statistics of the combined process are Poisson. This is an important result.

4.10.3 Multiplier dispersion

A multiplier with k stages is another example of cascaded operation. To obtain the variance at the k^{th} stage, for a single photoelectron released at d_1 , we extend the method adopted to arrive at (4.44). The output of the first stage serves as the input to the second stage and so forth. Dynodes with different mean gains and variances are considered in the analysis that follows, but without reference to any particular statistical distribution. For the first stage, the mean and variance are δ_1 and σ_1^2 , respectively. For the variance at the second stage, var_2 , we use the cascade formula given by (4.44):

$$\begin{aligned}
 m_1 &= \delta_1, \\
 \text{var}_1 &= \sigma_1^2; \\
 m_2 &= \delta_1 \delta_2, \\
 \text{var}_2 &= (\sigma_1^2 \delta_2^2 + \sigma_2^2 \delta_1); \\
 m_3 &= \delta_1 \delta_2 \delta_3, \\
 \text{var}_3 &= (\sigma_1^2 \delta_2^2 + \sigma_2^2 \delta_1) \delta_3^2 + \sigma_3^2 \delta_1 \delta_2 \\
 &\quad = (\sigma_1^2 \delta_2^2 \delta_3^2 + \sigma_2^2 \delta_1 \delta_3^2 + \sigma_3^2 \delta_1 \delta_2) \\
 &= (\delta_1 \delta_2 \delta_3)^2 \left(\frac{\sigma_1^2 \delta_2^2 \delta_3^2}{(\delta_1 \delta_2 \delta_3)^2} + \frac{\sigma_2^2 \delta_1 \delta_3^2}{(\delta_1 \delta_2 \delta_3)^2} + \frac{\sigma_3^2 \delta_1 \delta_2}{(\delta_1 \delta_2 \delta_3)^2} \right) \quad (4.46) \\
 &= (\delta_1 \delta_2 \delta_3)^2 \left(\frac{\sigma_1^2}{\delta_1^2} + \frac{\sigma_2^2}{\delta_1 \delta_2^2} + \frac{\sigma_3^2}{\delta_1 \delta_2 \delta_3^2} \right); \\
 \text{var}_k &= m_k^2 \left(\frac{\sigma_1^2}{\delta_1^2} + \frac{\sigma_2^2}{\delta_1 \delta_2^2} + \frac{\sigma_3^2}{\delta_1 \delta_2 \delta_3^2} + \cdots + \frac{\sigma_k^2}{\delta_1 \delta_2 \delta_3 \cdots \delta_{k-1} \delta_k^2} \right),
 \end{aligned}$$

with an overall mean gain of $m_k = (\delta_1 \times \delta_2 \times \delta_3 \times \cdots \times \delta_k)$. Note that the first dynode makes the major contribution to the total variance, while contributions from successive dynodes diminish as the anode is approached. It is also clear that the gain of the first stage should be uniform and exceed all the other stage gains to achieve optimal resolution. Consider a special case of (4.46), described by Poisson statistics, by assuming identical gain and variance for all stages:

$$\begin{aligned}
 \text{var}_k &= \delta^{2k} \left(\frac{1}{\delta} + \frac{1}{\delta^2} + \frac{1}{\delta^3} \cdots + \frac{1}{\delta^k} \right) \\
 &= \frac{\delta^{2k}}{\delta} \left(1 + \frac{1}{\delta} + \frac{1}{\delta^2} \cdots + \frac{1}{\delta^{k-1}} \right) \quad (4.47)
 \end{aligned}$$

This series within the brackets sums to $\delta / (\delta - 1)$ and we have

$$\frac{\text{var}_k}{g^2} = \frac{1}{\delta} \frac{\delta}{(\delta - 1)}. \quad (4.48)$$

$$\frac{\text{var}_k}{\langle g \rangle^2} = \frac{1}{(\delta - 1)}. \quad (4.49)$$

$$1 + \frac{\text{var}(g)}{\langle g \rangle^2} = \frac{\delta}{(\delta - 1)} = ENF. \quad (4.50)$$

The excess noise factor ENF follows from the assumption of Poisson statistics at all gain stages; it is a measure of the degradation in resolution caused by an electron multiplier purely on statistical grounds. Multipliers, other than SiPM types, do not conform to such statistics but generate pulse height distributions with an excess of small pulses, such as those exhibited in (b) of Fig. 4.2. Nevertheless, the theoretical treatment points to several important generalities. The form of (4.47) indicates that the contribution to noise from each additional dynode decreases as $1/\delta$ from its predecessor and hence the importance of a high-gain first dynode. The magnitude of the gain of successive dynodes is relatively unimportant where noise is concerned, although uniformity over the entire dynode collection area is important, especially for the first two or three stages. Venetian blind dynodes are arguably the poorest for uniformity because of geometrical shadowing (see Chapter 5, Figs 5.20 and 5.21). Note that ENF can be computed from any measured or assumed SER, without knowledge of the mean gain, $\langle g \rangle$. This is because $\text{var}(g)/\langle g \rangle^2$ is dimensionless with customary units for the abscissa in terms of pulse height or photoelectrons equivalent. The relative variance for a single-stage multiplier is $1/\delta$, which represents the best resolution attainable. Successive gain stages add more noise, but never reduce it. Optimal noise performance is governed by the distribution at the first dynode, and we note that (4.49) approaches $1/\delta$ for $\delta \gg 1$. If we choose a PMT with high-gain dynodes, with say $\delta \approx 10$, according to this statistical model, we can derive additional gain from the multiplier without adding significant noise to the signal. A PMT with high-gain dynodes will approach this performance with a relative variance of $1/\delta$. This follows from (4.46) by assuming σ_1 and δ_1 for the first stage, and a lower δ for each of the remainder:

$$\begin{aligned} \text{var}_k &= (\delta_1 \delta^{k-1})^2 \left(\frac{1}{\delta_1} + \frac{\delta}{\delta_1 \delta^2} + \frac{\delta}{\delta_1 \delta^3} + \cdots + \frac{\delta}{\delta_1 \delta^{k-1}} \right) \\ &= \frac{(\delta_1 \delta^{k-1})^2}{\delta_1} \left(1 + \frac{1}{\delta} + \frac{1}{\delta^2} + \cdots + \frac{1}{\delta^{k-1}} \right); \\ \frac{\text{var}_k}{\langle g \rangle^2} &= \frac{\delta}{\delta_1} \left(\frac{1}{\delta - 1} \right). \\ &= \frac{\delta}{\delta_1(\delta - 1)} \end{aligned} \quad (4.51)$$

If $\delta_1 = \delta$, (4.51) reverts to (4.49).

The relationships (4.46) and (4.51) are central to the design and application of PMTs; note that the relative variance is reduced by a factor of δ/δ_1 by comparing

(4.49) and (4.51). Some 50 years ago, RCA developed a PMT with a high-gain first dynode and with the secondary emission characteristic shown in Chapter 5, Fig. 5.11. GaP provides gain of up to 100 at a sufficiently high accelerating voltage with V_{k-d1} of the order of 500 V. This high gain should be compared with BeO and K₂SbCs with respective yields of about 5 and 15, only. Also GaP PMTs generally have 12 dynodes, but only the first one is GaP, with the remainder being BeO. Designers are thus able to take full advantage of the predictions of (4.51). Manufacturers offer 12-stage PMTs with BeO dynodes, with the exception of a high-gain K₂SbCs first dynode—the same material as the photocathode. Nowadays, most PMTs include high-gain dynodes throughout, or at least a high-gain first dynode. BeO dynodes offer superior performance with regard to pulse height linearity at the upper end of the signal scale; the combination of a high-gain first stage with otherwise BeO dynodes gives both good resolution and wide dynamic range in pulse height. A high-gain first dynode leads to a well-resolved SER, of particular importance in photon counting. If all pulses originating as single photoelectrons have a narrow spread in pulse heights, Δh , this gives scope for eliminating background pulses, both undersized and oversized that lie outside $h \pm \Delta h$. This in turn leads to better signal-to-noise and signal-to-background performance. Resolution on multi-photoelectron pulses, as in scintillation spectroscopy, depends on high cathode efficiency and on the quality of the SER; this is investigated in 4.14.

4.10.4 Pdf at the k^{th} dynode

The statistical methods so far employed generate the mean and variance at every stage of a multiplier, without revealing the corresponding pdfs. Lombard and Martin (1961) describe a way of generating pdfs for an ideal multiplier with all stages of identical gain. The generating function for the k^{th} stage is derived from the previous one; thus,

$$G_k(u) = G_1[G_{k-1}(u)]. \quad (4.52)$$

The generating function at the k^{th} stage is

$$\begin{aligned} G_k(u) &= \sum_{n=0}^{\infty} P_k(n) u^n; \\ \frac{dG_k(u)}{du} &= 0 + P_k(1) + 2P_k(2)u + 3P_k(3)u^2 + \dots; \\ \frac{dG_k(u)}{du}_{u=0} &= P_k(1); \\ \frac{d^2 G_k(u)}{du^2}_{u=0} &= 2P_k(2); \\ \frac{d^n G_k(u)}{du^n}_{u=0} &= n!P_k(n). \end{aligned}$$

Rearranging we have for the n^{th} derivative

$$P_k(n) = \frac{1}{n!} \frac{d^n}{du^n} G_k(u)_{u=0}. \quad (4.53)$$

Equation (4.53) reveals the relationship between the pulse height distribution, $P_k(n)$, and its generating function. From (4.52), we have

$$\begin{aligned} \frac{d^n}{du^n} G_k(u) &= \frac{d^{n-1}}{du^{n-1}} \frac{d}{du} (G_1[G_{k-1}(u)]) \\ &= \frac{d^{n-1}}{du^{n-1}} \frac{d}{du} (e^{-\delta} \exp(\delta G_{k-1}(u))) \\ &= \frac{d^{n-1}}{du^{n-1}} \{e^{-\delta} \exp(\delta G_{k-1}(u)) \times \delta G'_{k-1}(u)\}. \end{aligned}$$

4

The expression $e^{-\delta} \exp(\delta G_{k-1}(u))$ is just $G_1(G_{k-1}(u))$; so, finally,

$$\begin{aligned} \frac{d^n}{du^n} G_k(u) &= \delta \frac{d^{n-1}}{du^{n-1}} (G_1[G_{k-1}(u)] \times G'_{k-1}(u)) \\ &= \delta \frac{d^{n-1}}{du^{n-1}} (G_k(u) \times G'_{k-1}(u)), \quad n \neq 0. \end{aligned} \quad (4.54)$$

The Leibnitz rule for differentiating the product of two functions f and g is

$$\frac{d^n}{du^n} (f \times g) = \sum_{i=0}^{n-1} \frac{n!}{i!(n-i)!} f^i \times g^{n-i},$$

where the superscripts denote differentiation:

$$\begin{aligned} \frac{d^n}{du^n} (G_k(u) \times G'_{k-1}(u)) &= \sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-i-1)!} \times G_k^i(u) \times G_{k-1}^{n-i-1}(u), \\ \frac{d^n}{du^n} G_k(u) &= \delta \times \sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-i-1)!} \times G_k^i(u) \times G_{k-1}^{n-i}(u). \end{aligned}$$

From (4.53), we have

$$\begin{aligned} P_k(n) &= \frac{\delta}{n!} \times \sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-i-1)!} \times i! P_k(i) \times (n-i)! \times P_{k-1}(n-i) \\ &= \frac{\delta}{n} \times \sum_{i=0}^{n-1} (n-i) \times P_k(i) \times P_{k-1}(n-i), \quad n \neq 0, k \geq 1. \end{aligned} \quad (4.55)$$

The case for $n = 0$ is treated separately. From (4.52) and (4.53),

$$G_k(0) = P_k(0) = G_1[G_{k-1}(0)] = P_1[P_{k-1}(0)] = e^{-\delta} \exp[\delta \times P_{k-1}(0)]. \quad (4.56)$$

The probability of recording a zero at the first dynode follows directly from the Poisson distribution as $e^{-\delta}$, and from the generating function:

$$P_1(0) = G_1(0) = e^{-\delta}e^0 = e^{-\delta}$$

and, for the second and third stages,

$$P_2(0) = G_2(0) = e^{-\delta} \exp(\delta \times P_1(0)) = e^{-\delta} \exp(\delta \times e^{-\delta});$$

$$P_3(0) = G_3(0) = e^{-\delta} \exp(\delta \times P_2(0)).$$

Thus, $P_k(n)$ can be derived for an initiating Poisson distribution, $P_1(n)$, by using the iterative relations (4.55) and (4.56). Pulse height distributions at the k^{th} dynode are given by equation (4.55) and are shown for $k = 1, 2$, and 3 in Fig. 4.6 for $\delta = 2$, and in Fig. 4.7 for $\delta = 5$. We note in Fig. 4.6 that the pulse height distribution broadens as the cascade propagates to subsequent dynodes, whereas in Fig. 4.7 there is little further dispersion beyond the first dynode—a direct result of the higher δ value assumed. Note that the abscissa has been scaled for $k = 2$ and for $k = 3$ by a factor, $\delta^k P_k(n/\delta^k)$. Superimposing distributions in this way highlights the relatively minor change in shape with increasing k . It is sufficient to look no further than the third stage of multiplication to achieve an accurate indication of the distribution at the anode. Figure 4.8 shows the rapidity with which the SER improves with increasing δ .

4

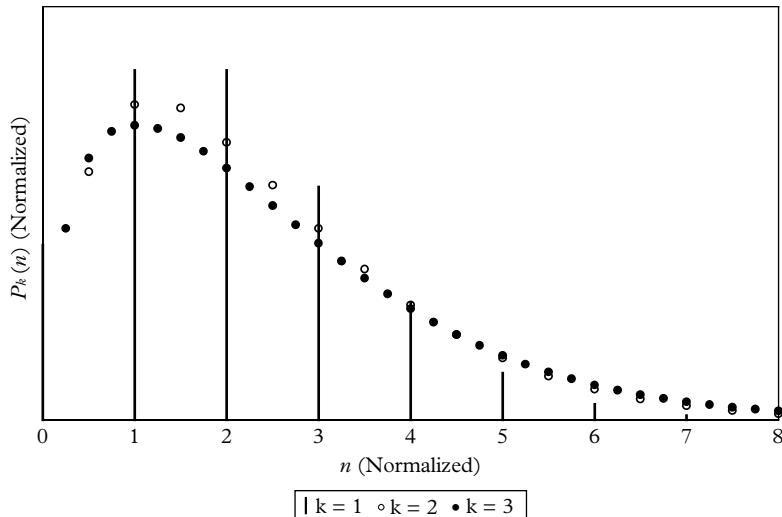


Fig. 4.6. Pulse height distributions, for single-electron input, at the first, second, and third dynodes of a multiplier with a low mean stage gain of $\delta = 2$. The vertical lines represent the Poisson distribution for the first stage of gain. Note how the density of points increase with k ; $P_1(0)$ is predictable and part of the distribution, although it is always absent from a measured spectrum—a zero cannot be recorded!

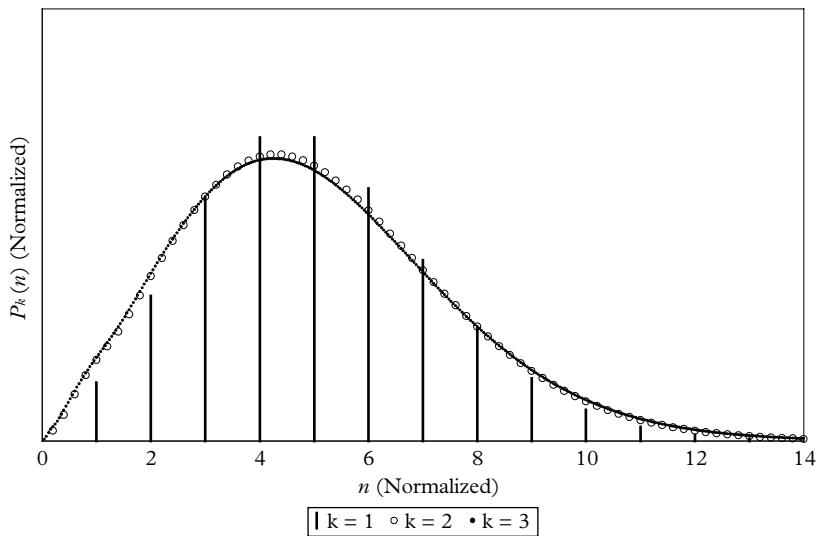


Fig. 4.7. Pulse height distributions, for single-electron input, at the first, second, and third dynodes of a multiplier with a mean stage gain of $\delta = 5$. These plots illustrate the predictions of 4.10.4 concerning the relative insignificance of the gain stages beyond d_1 .

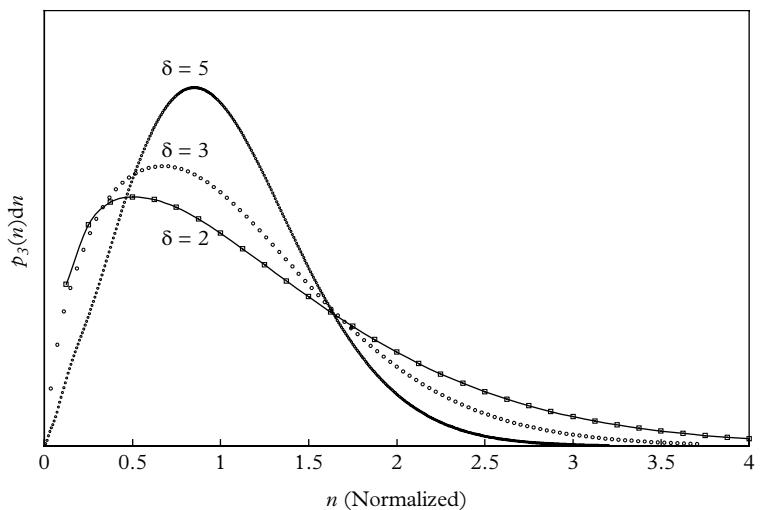


Fig. 4.8. Illustrating how the width of the pulse height distribution at the third dynode narrows with increasing δ .

The probability that a photoelectron fails to propagate is calculated from (4.56) with the results given in Table 4.3 for $\delta = 2$ to 5. We note for $\delta = 2$ that 13.5 % of photoelectrons fail to propagate beyond the first dynode; 17.7 % fail to propagate beyond the second dynode, and so on. For $\delta > 3$ the losses at the first dynode are only a few per cent, and further reductions at subsequent dynodes are very small. For a secondary emission process obeying Poisson statistics, we note that the effect of truncated propagation is equivalent to a reduction in collection efficiency. The losses are more serious than the above purely statistical treatment predicts because of losses incurred from electron scattering off the dynodes.

The practical value of this elegant statistical treatment might be questionable. It is useful, however, because it draws attention to the important role of a multiplier with high stage gain, and in addition, one that conforms to Poisson statistics. The prediction that a fraction of single photoelectrons fail to propagate is equivalent to a loss in collection efficiency, although the physics of collection efficiency is complicated and contentious, as will become clear in 10.2. No doubt, terminated propagation is a physical reality but reliable corrections to collection efficiency cannot be made based on the theoretical $P_k(0)$ values.

Table 4.3 Probability $P_k(0)$ that a photoelectron fails to propagate beyond the first, second, or third dynode.

δ	$P_1(0)$	$P_2(0)$	$P_3(0)$
2	0.1353	0.1774	0.1930
3	0.0498	0.0578	0.0592
4	0.0183	0.0197	0.0198
5	0.0067	0.0070	0.0070

4.11 Shot noise in DC detection

The discrete nature of electrical current is governed by the Schoty noise formulation. Consider a steady flux of light producing a mean photocurrent of I_0 . Where light levels are low, I_0 is too small to measure directly and as already discussed, an electron multiplier provides the best means of charge amplification. The mean anode current for a k-stage multiplier is $I_a = \delta^k I_0$ but subject to fluctuation from the combination of two noise sources: first originating in the photocathode, and second from the electron multiplier. The fluctuation in photocathode current is given by the well-known formula for noise,

$$\langle i_0^2 \rangle^{1/2} = (2eI_0\Delta f)^{1/2}, \quad (4.57)$$

with $\langle i_0^2 \rangle^{1/2}$ referred to as the root mean square noise (rms), and e being the electronic charge; $I_0 = me$ derives from a flow of m electrons per unit time. The signal-to-noise ratio for shot noise, expressed in terms of m , is

$$(S/N)^2 = \frac{I_0^2}{2eI_0\Delta f} = \frac{me}{2e\Delta f} = \frac{m}{2\Delta f}. \quad (4.58)$$

But I_0 stems from m counts per second, with an accompanying fluctuation in accordance with Poisson statistics. In terms of counts:

$$(S/N)^2 = \frac{m^2}{m} = m. \quad (4.59)$$

The expressions in (4.58) and (4.59) describe the same physical phenomenon and therefore must be equivalent; this follows if Δf is taken as 0.5 Hz. Measuring time is related to bandwidth by a Fourier transform so the equivalence of these formulations is expected.

The expression in (4.57) represents the noise associated with the photocathode but the noise at the anode, induced by the multiplier, is equally important. The mean current, I_1 , leaving the first dynode is given by δI_0 , and its fluctuation derives from the combined action of I_0 and I_1 adding in quadrature, with

$$\langle i_1^2 \rangle = \delta(2eI_0\Delta f) + \delta^2(2eI_0\Delta f).$$

Similarly, the fluctuation at the second dynode is the combined action of I_0 , I_1 , and I_2 , leading to the formulation at the k^{th} dynode, $\langle i_k^2 \rangle$, given below:

$$\langle i_0^2 \rangle = 2eI_0\Delta f,$$

$$\langle i_1^2 \rangle = \langle i_0^2 \rangle \delta(1 + \delta),$$

$$\langle i_2^2 \rangle = \langle i_0^2 \rangle \delta^2(1 + \delta + \delta^2),$$

$$\langle i_k^2 \rangle = \langle i_0^2 \rangle \delta^k(1 + \delta + \delta^2 + \dots + \delta^k).$$

The series sums as follows:

$$1 + \delta + \delta^2 + \dots + \delta^k = \frac{\delta^{k+1} - 1}{\delta - 1},$$

leading to

$$\langle i_k^2 \rangle = \langle i_0^2 \rangle \delta^k \frac{\delta^{k+1} - 1}{\delta - 1}. \quad (4.60)$$

The signal-to-noise ratio at the input and output are, respectively,

$$(S/N)_{\text{ip}}^2 = \frac{I_0^2}{\langle i_0^2 \rangle},$$

$$(S/N)_{\text{op}}^2 = \frac{(\delta^k I_0)^2}{\langle i_0^2 \rangle} \times \frac{\delta - 1}{\delta^k (\delta^{k+1} - 1)}.$$

Since $\delta^{k+1} \gg 1$,

$$\frac{(S/N)_{\text{ip}}^2}{(S/N)_{\text{op}}^2} = \frac{\delta^{k+1} - 1}{\delta^k (\delta - 1)} = \frac{\delta}{\delta - 1}$$

or

$$\frac{(S/N)_{\text{op}}^2}{(S/N)_{\text{ip}}^2} = \frac{\delta - 1}{\delta}. \quad (4.61)$$

The signal-to-noise ratio at the photocathode is optimal while, at the anode, the ratio is reduced by a small fraction, provided that the stage gain is sufficiently large. From (4.60),

$$\langle i_k^2 \rangle = \langle i_0^2 \rangle \delta^{2k} \left(\frac{\delta}{\delta - 1} \right)$$

and the relative variance at the anode, $\text{var}_k / \langle g \rangle^2$, is

$$\frac{\langle i_k^2 \rangle}{\delta^{2k}} = \langle i_0^2 \rangle \left(\frac{\delta}{\delta - 1} \right).$$

We note that the ENF $\delta / (\delta - 1)$ is the same as the expression introduced in (4.50) in 4.10.3, as it must be.

4.12 Noise in multi-photon excitation

Recall that (4.47) refers to the statistical spread in the output of an electron multiplier excited by single photoelectrons. Equally important is knowledge of the dispersion at the anode for pulses initiated by n_{pe} photoelectrons, Poisson distributed. The derivation starts with equation (4.44):

$$\text{var}_{AB} = (\sigma_A)^2 m_B^2 + (\sigma_B)^2 m_A,$$

where A is taken as the contribution from multi-photoelectron statistics, and B the noise from the multiplier. If the QE is η , the mean number of photoelectrons is $\eta n_p = n_{pe}$, and the variance of the multiplier is $\sigma_{mk}^2 = \delta^{2k}/(\delta - 1)$ together with

$$\begin{aligned} (\sigma_A)^2 &= \eta n_p, \\ m_B^2 &= \delta^{2k} = \langle g \rangle^2, \\ (\sigma_B)^2 &= \sigma_{mk}^2 = \delta^{2k} \left(\frac{1}{\delta - 1} \right) = \frac{\langle g \rangle^2}{\delta - 1}, \\ m_A &= \eta n_p. \end{aligned}$$

Substituting into (4.44), the noise at the anode, σ_{op} , is

$$\begin{aligned} (\sigma_{op})^2 &= \eta n_p \delta^{2k} + \delta^{2k} \left(\frac{1}{(\delta - 1)} \right) \eta n_p \\ &= \eta n_p \delta^{2k} \left(1 + \frac{1}{\delta - 1} \right) \\ &= \eta n_p \delta^{2k} \left(\frac{\delta}{\delta - 1} \right). \end{aligned}$$

Therefore,

(4.62)

$$\begin{aligned}
 (S/N)_{\text{op}} &= \frac{\eta n_p \delta^k}{\left[\eta n_p \delta^{2k} \left(\frac{\delta}{\delta - 1} \right) \right]^{1/2}} \\
 &= \frac{(\eta n_p)^{1/2}}{\left(\frac{\delta}{\delta - 1} \right)^{1/2}}.
 \end{aligned}$$

4

We can generalize (4.62), as will be shown in 4.13, to derive the signal-to-noise ratio for any analytical or experimental distribution; thus,

$$S/N = \frac{(\eta n_p)^{1/2}}{\left(1 + \frac{\text{var}(g)}{\langle g \rangle^2} \right)^{1/2}}.$$

It is important to remember that (4.62) refers to an ideal multiplier for which all gain stages are identical and equal to δ . It is left as an exercise to show that the variance at the anode is a factor of 2 higher for exponential dynode statistics. The inverse of $(S/N)_{\text{op}}$ with inclusion of the 2.35 factor gives the resolution, R , in terms of fwhm per peak height of the distribution:

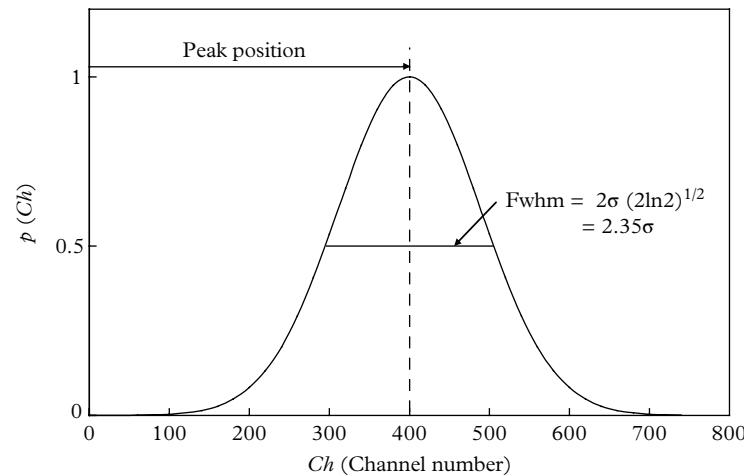
$$R = \frac{2.35}{(\eta n_p)^{1/2}} \times \left(\frac{\delta}{\delta - 1} \right)^{1/2}, \quad (4.63)$$

and (4.63) becomes

$$n_{\text{pe}} = \frac{ENF}{\left(\frac{R}{2.35} \right)^2}. \quad (4.64)$$

Resolution varies as $(\eta n_p)^{-1/2}$ once δ is sufficiently large. If we take $\delta = 10$, easily achievable for the first dynode, then $ENF = 1.11$, and the effect on resolution varies as $(ENF)^{1/2}$, which is ~ 1.06 in this instance and so not very significant. In practice, backscattering, non-uniformity of δ over the first dynode area, and variable collection from d_1 and d_2 result in ENF values between 1.2 and 1.5 for PMTs but close to unity for SiPMs.

PMT manufacturers provide comprehensive test ticket data with every PMT. However, this information cannot answer a fundamental question that often arises when users are setting up test equipment for spectroscopy: how many photons or photoelectrons correspond to a particular peak in a spectrum? Equation (4.64) suggests an efficient method for calibrating light signals incident on a PMT. A pulsed light source, such as a scintillator, generally obeys Poisson statistics, and (4.64) reveals that the dispersion of a multi-photoelectron pulse height distribution (MPHD) should, subject to certain conditions, lead to the mean number of photoelectrons per pulse. If the QE, η , is known, or can be guessed, this leads to n_p . The normal distribution of Fig. 4.1 is reproduced in Fig. 4.9, with the abscissa expressed in arbitrary MCA numbers.



4

Fig. 4.9. The relationship, for a normal distribution, between the resolution of a pulse height spectrum and σ . Determination of the mean number of photoelectrons in a pulsed light source can be deduced from such a spectrum.

The aim is to devise a scale in photoelectrons in preference to channels. In spectroscopy, the resolution, R , is defined as the fwhm peak height divided by the peak position of the distribution. For the distribution shown in Fig. 4.9, we have

$$R \approx \frac{500 - 300}{400} = 0.5 \text{ or, equivalently, } 50\%,$$

$$n_{pe} = \left(\frac{2.35}{R} \right)^2 \approx 22 \text{ pe},$$

which is consistent with Fig. 4.1. The estimate of the number of photoelectrons should include the noise factor (if known), although its contribution is usually small. It is clear from the similarity of the three distributions that the method will work reasonably well for n_{pe} in excess of about 10.

The second calibration method is based on knowing the spectrum of the SER for a given PMT, operated obviously under the same conditions. The channel corresponding to the SER peak is taken as 1 photoelectron equivalent, from which the abscissa calibration follows. Strictly, the mean of the SER should be used, although for a quality PMT the two parameters lie close together. This method of calibration does not suffer the same limitation as the previous one since the SER and the multi-photoelectron distributions are subject to the same multiplier statistics. Nevertheless, the pulse height resolution serves as a quick and easy means to rough calibration in terms of photoelectrons and photons. The author's experience is that the estimates provided by such calibration methods should be reliable to better than 20 %.

4.13 Continuous SER distributions

To illustrate the use of generating functions for situations where the variables may be a mixture of discrete or continuous, we further explore the subject of PMT resolution. The question is, given a pulsed source of light, with the discrete multi-photoelectron spectrum shown in Fig. 4.10(a), and a PMT with an SER shown in Fig. 4.10(b), what are the PMT output pulse height distribution, variance, and mean? No assumptions are made initially about the statistical distributions; they could be discrete, such as Poisson and binomial, or continuous functions, either analytical or experimental.

If we have m pe in a pulse, with a distribution, $p(m)$ given in Fig. 4.10(a), each of which produces a charge q_1, q_2, \dots, q_m at the anode, picked randomly in accordance with the SER of Fig. 4.10(b), then the charge at the anode, Q , is

$$Q = \sum_{i=1}^m q_i. \quad (4.65)$$

The probability distribution for Q has the form given in Fig. 4.10(c), described by

$$p(Q) = \sum_{m=1}^{\infty} p(m) p \times \left(\sum_{i=1}^m q_i = Q \right). \quad (4.66)$$

Therefore,

$$p(Q) = p(1) p(q_1 = Q) + p(2) p(q_1 + q_2 = Q) + p(3) p(q_1 + q_2 + q_3 = Q) + \dots \quad (4.67)$$

Equation (4.67) illustrates the manifold ways in which a charge Q is generated at the anode from an initial combination of m pe; these m pe are individually acted upon by the multiplier to generate anode charges q_1, q_2, \dots, q_m .

Any one of the forms given in Table 4.2 provides a suitable generating function for the SER. Using, for example, $G_q(\nu)$, we have

$$\begin{aligned} G_q(\nu) &= \int e^{\nu q} \times p(q) dq \\ &= \int \left(1 + \nu q + \frac{(\nu q)^2}{2!} + \frac{(\nu q)^3}{3!} + \dots \right) \times p(q) dq \end{aligned} \quad (4.68)$$

We see immediately that $G_q(0) = 1$, assuming that the distribution $p(q)$ is normalized:

$$G'_q(\nu) = \int \left(q + \nu q^2 + \frac{\nu^2 q^3}{2!} + \dots \right) \times p(q) dq, \quad (4.69)$$

$$G'_q(0) = \int p(q) dq = q_0, \quad (4.70)$$

$$G''_q(\nu) = \int (q^2 + \nu q^3 + \dots) \times p(q) dq,$$

$$G''_q(0) = \int p(q) q^2 dq = \langle q^2 \rangle. \quad (4.71)$$

From (4.38) we have, for the mean and variance of the SER,

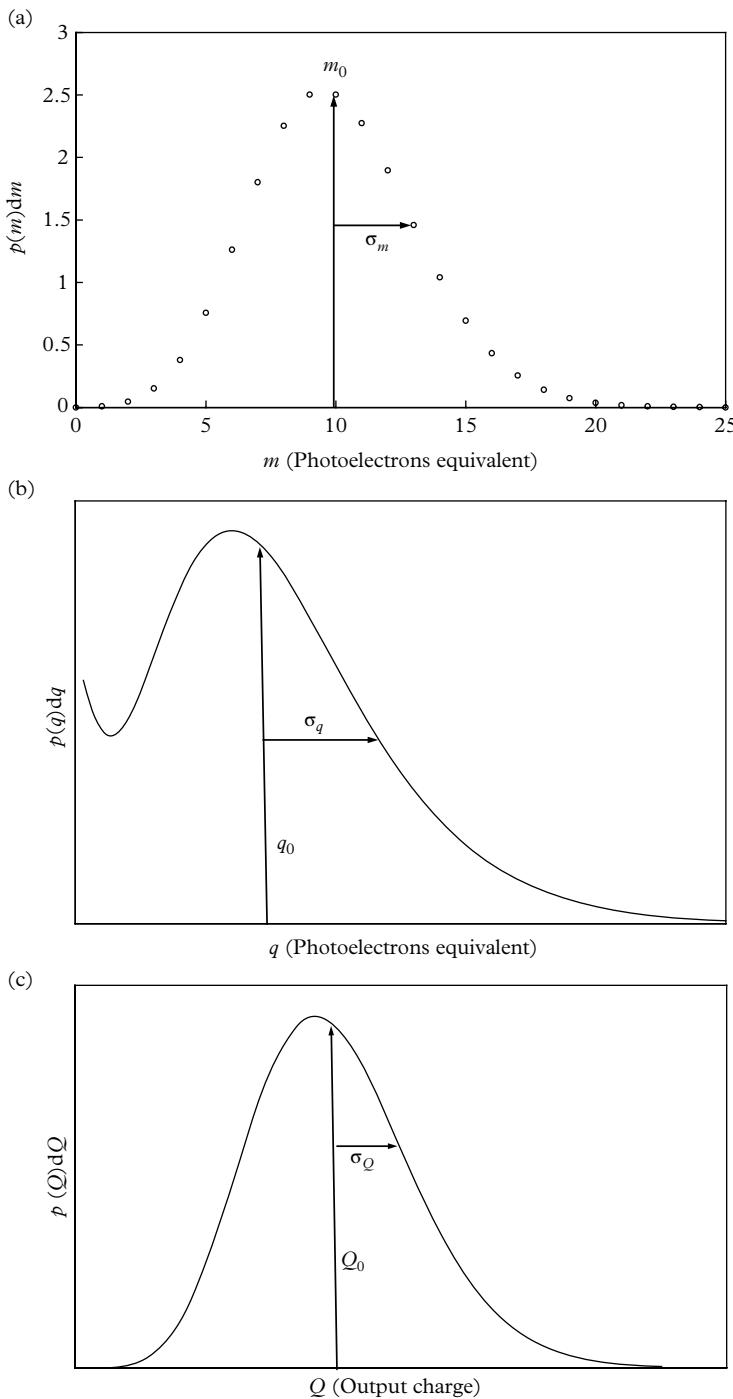


Fig. 4.10. (a) A pulsed light source generates a discrete spectrum of photoelectrons characterized by a mean m_0 and standard deviation $\sqrt{m_0}$. (b) The SER is assumed to be continuous. (c) The PMT output for the combined processes (a) and (b).

$$\begin{aligned} m &= G_q'(0) = q_0, \\ \text{var}_q &= G_q''(0) - (G_q'(0))^2 = \langle q^2 \rangle - q_0^2. \end{aligned} \quad (4.72)$$

In 4.7 we derived the generating function for independent statistical processes whose outcomes were summed; for m events acted upon by the same distribution (in this case, the SER), the generating function is

$$\begin{aligned} G_m(\nu) &= (G_q(\nu))^m, \\ G_Q(\nu) &= \sum_{m=1}^{\infty} p(m) \times (G_q(\nu))^m \end{aligned} \quad (4.73)$$

and

$$\begin{aligned} G_Q'(\nu) &= \sum_{m=1}^{\infty} p(m)m \times (G_q(\nu))^{m-1} \times G_q'(\nu) \\ &= G_q'(\nu) \sum_{m=1}^{\infty} p(m)m \times (G_q(\nu))^{m-1}; \end{aligned} \quad (4.74)$$

since $G_q'(0) = q_0$, the mean pulse height of the SER, and $G_q(0) = 1$,

$$G_Q'(0) = q_0 \sum_{m=1}^{\infty} p(m)m = q_0 m_0 = Q_0, \quad (4.75)$$

an obvious result, where Q_0 is the mean pulse height at the output, and m_0 the mean of the photoelectron distribution. Differentiating (4.74), we have

$$G_Q''(\nu) = \sum_{m=1}^{\infty} p(m)m \times ((m-1) \times (G_q(\nu))^{m-2} \times (G_q'(\nu))^2 + (G_q(\nu))^{m-1} \times G_q''(\nu)).$$

Using $G_q(0) = 1$, $G_q'(0) = q_0$, and $G_q''(0) = \langle q^2 \rangle$ together with (4.72) in the above equation we obtain

$$\begin{aligned} G_Q''(0) &= \sum_m p(m)m \times ((m-1)q_0^2 + \langle q^2 \rangle) \\ &= \sum_m p(m)m^2 q_0^2 - \sum_m p(m)m q_0^2 + \sum_m p(m)m(\text{var}_q + q_0^2) \\ &= \langle m^2 \rangle q_0^2 + m_0 \text{var}_q \\ &= \text{var}_m q_0^2 + m_0^2 q_0^2 + m_0 \text{var}_q. \end{aligned}$$

By using (4.7) and, analogously with (4.38), noting $m_0^2 q_0^2 = Q_0^2$, we find that

$$\begin{aligned} G_Q''(0) &= \text{var}_Q + Q_0^2, \\ \text{var}_Q + Q_0^2 &= \text{var}_m q_0^2 + Q_0^2 + m_0 \text{var}_q, \\ \frac{\text{var}_Q}{Q_0^2} &= \frac{\text{var}_m}{m_0^2} + \frac{\text{var}_q}{m_0 q_0^2}. \end{aligned} \quad (4.76)$$

If $p(m)$ is Poisson, for example, then $\text{var}_m = m_0$, and

$$\frac{\text{var}_Q}{Q_0^2} = \left(\frac{1}{m_0} \right) \times \left(1 + \frac{\text{var}_q}{q_0^2} \right). \quad (4.77)$$

As we have seen before, the actual distributions of the variables is not required—only their mean and variance. Equation (4.77) is given by Jones, Oliver, and Pike (1971). In their derivation, Q is considered the integration of a train of output pulses over a time, T . If m_0 is the mean rate per second, then (4.77) becomes

$$\frac{\text{var}_Q}{Q_0^2} = \left(\frac{1}{m_0 T} \right) \times \left(1 + \frac{\text{var}_q}{q_0^2} \right). \quad (4.78)$$

Equation (4.78) describes the noise that accompanies DC detection from a source of light described by Poisson statistics. We note in both equations that the resolution is degraded by the factor $(1 + \text{var}_q/q_0^2)$, arising from the noise in the gain process; resolution also depends on $1/m_0$ in pulse encoding and on $1/(m_0 T)$ in DC measurements. To obtain the best resolution, we should select a PMT with a well-resolved SER and with a high QE η (since m_0 is proportional to QE).

4.14 Multi-photoelectron pdfs

The first MPHD reported in the literature is the one due to Morton *et al.* (1968) at RCA. Recognizing the importance of high and uniform d_1 gain, RCA developed the renowned 8850 PMT with a GaP first dynode and an achievable stage gain of up to 100. The high and uniform gain of the first dynode revealed the imprint of up to 5 pe, derived from a suitably adjusted multi-photoelectron source (see 4.1.1). Further detectors capable of even higher resolution have been developed since that time. These include MCPMTs, HPDs, and, more recently, several different types of SiPMs based on multi-pixel APDs. The last two mentioned are capable of resolving up to 20 peaks. Detailed descriptions of all these devices and their performance capabilities are found in the conference proceedings of the International Conference on New Developments in Photodetection, Beaune, France, for 1999, 2002, and 2005 (Barletta *et al.* 2000, 2003, 2006). In practical terms, a PMT with an SER peak-to-valley ratio in excess of 3 and a resolution better than 60 % is capable of uncovering a structured multi-electron spectrum. Despite the obvious importance of the MPHD in a wide range of detectors, there is, to the author's knowledge, no published statistical treatment of the subject that is wholly satisfactory. The analysis included in (4.79) to (4.86) that follows is taken from Wright (2007).

The starting point is usually the assumption of a Gaussian distribution or a combination of an exponential and a Gaussian for the SER—this is of doubtful merit. However, for completeness in the treatment that follows, we initially choose Poisson statistics for the multiplier, followed by the general case that permits the

use of any mathematical distribution, or, more importantly, an experimentally determined SER. It was shown in (4.51) that the first stage of a multiplier is most important in determining resolution where the first stage is of high gain (see Fig. 4.7). Since we are dealing with devices of high d_1 gain, we can use an analytical model with just a single high-gain stage of multiplication for the initial investigation. As before, the two distributions A and B are cascaded; A describes the light detection statistics, for which Poisson statistics are justifiably assumed, while B refers to the statistics of the multiplier, which initially is unspecified. In addition, A occurs first and B operates on the outcome of A. We derive an expression for the combined statistics for any distribution representing B. Given the SER of a photon-detecting device, whether it is theoretical or experimentally determined, we are able to predict the MPHD for all detectors.

The probability distributions for A and B are represented by $p(n)$ and $q(s)$, respectively. The generating function, $G_B(u)$, for the dynodes is

$$G_B(u) = q(0) + q(1)u + q(2)u^2 + \dots + q(s)u^s + \dots, \quad (4.79)$$

where s may be infinite, such as a Poisson distribution, or finite, with the number of terms, $s + 1$, taken to represent a set of experimental probabilities describing the SER. The generating function for the cascaded process is

$$\begin{aligned} G_{AB}(u) &= G_A[G_B(u)] \\ &= p(0) + p(1)G_B(u) + p(2)G_B(u)^2 + p(3)G_B(u)^3 + \dots \end{aligned} \quad (4.80)$$

The function $G_{AB}(u)$ may also be written in terms of the cascaded probability distribution $P(r)$, which is the required output distribution:

$$G_{AB}(u) = P(0) + P(1)u + P(2)u^2 + P(3)u^3 + \dots \quad (4.81)$$

Differentiating (4.79) and (4.81) r times, denoted by a superscript in parenthesis, and setting u to 0, we have

$$\begin{aligned} G_B^{(r)}(0) &= r!q(r), \\ G_{AB}^{(r)}(0) &= r!P(r), \end{aligned} \quad (4.82)$$

where, for presentation, we define $G_{AB}^{(0)}(u)$ to be the undifferentiated function $G_{AB}(u)$. We substitute the $p(n)$ terms in (4.80) by Poisson probabilities to describe the light emission process, based on a mean of m_0 photoelectrons. This leads to the combined generating function, $G_{AB}(u)$:

$$\begin{aligned} G_{AB}(u) &= e^{-m_0} + e^{-m_0}m_0G_B(u) + \frac{e^{-m_0}}{2!}m_0^2\left(G_B(u)\right)^2 + \dots \\ &= e^{-m_0}\left(1 + m_0G_B(u) + m_0^2\left(\frac{G_B(u)}{2!}\right)^2 + \dots\right), \\ G_{AB}^{(0)}(u) &= e^{-m_0} \times \exp\left(m_0G_B(u)\right). \end{aligned}$$

From (4.79), $G_B(0) = q(0)$; consequently,

$$\begin{aligned}
G_{AB}^{(0)}(0) &= e^{-m_0} \times \exp(m_0 q(0)) \\
&\quad \exp(-m_0(1 - q(0)));
\end{aligned} \tag{4.83}$$

$$\begin{aligned}
G_{AB}^{(1)}(u) &= m_0 G_{AB}^{(0)}(u) \times G_B^{(1)}(u) \\
&= m_0(G_B^{(1)}(u) \times G_{AB}^{(0)}(u));
\end{aligned}$$

$$G_{AB}^{(2)}(u) = m_0(G_B^{(2)}(u) \times G_{AB}^{(0)}(u) + G_B^{(1)}(u) \times G_{AB}^{(1)}(u))$$

and

$$\begin{aligned}
G_{AB}^{(3)}(u) &= m_0(G_B^{(3)}(u) \times G_{AB}^{(0)}(u) + 2G_B^{(2)}(u) \\
&\quad \times G_{AB}^{(1)}(u) + G_B^{(1)}(u) \times G_{AB}^{(2)}(u)), \\
G_{AB}^{(r+1)}(u) &= m_0 \sum_{k=0}^r \frac{r!}{k!(r-k)!} G_B^{(r+1-k)}(u) \times G_{AB}^{(k)}(u).
\end{aligned} \tag{4.84}$$

From (4.82),

$$\begin{aligned}
G_{AB}^{(r+1)}(0) &= (r+1)! \times P(r+1), \\
G_B^{(r+1-k)}(0) &= (r+1-k)! \times q(r+1-k), \\
G_{AB}^{(k)}(0) &= k!P(k).
\end{aligned}$$

Substituting these three expressions into (4.84), we obtain

$$\begin{aligned}
(r+1)!P(r+1) &= m_0 \sum_{k=0}^r \frac{r!}{k!(r-k)!} G_B^{(r+1-k)}(0) \times G_{AB}^{(k)}(0), \\
P(r+1) &= m_0 \sum_{k=0}^r \frac{(r+1-k)! \times q(r+1-k)k! \times P(k)}{(r+1) \times k!(r-k)!} \\
&= \frac{m_0}{(r+1)} \sum_{k=0}^r (r+1-k) \times q(r+1-k) \times P(k).
\end{aligned} \tag{4.85}$$

From (4.81) and (4.83), we see that

$$P(0) = G_{AB}^{(0)}(0) = \exp(-m_0(1 - q(0))).$$

Knowing $P(0)$, we can calculate $P(1)$ from (4.85), which gives

$$\begin{aligned}
P(1) &= m_0 q(1) \times \exp(-m_0(1 - q(0))) \\
&= m_0 q(1) P(0).
\end{aligned}$$

Continuing this iteration yields the complete set of $P(r)$.

The analysis so far has been general since the multiplier statistics, $q(r)$, are unspecified. Now consider a case in which Poisson statistics apply to photodetection

and to the multiplier. It is assumed that only d_1 contributes to multiplier dispersion, that is,

$$q(r) = \exp(-\delta_1) \times \frac{\delta_1^r}{r!},$$

leading to

$$P(r+1) = \frac{m_0 \delta_1 \times \exp(-\delta_1)}{(r+1)} \times \sum_{k=0}^r \frac{(\delta_1)^{r-k} \times P(k)}{(r-k)!}. \quad (4.86)$$

Low-light-level detectors can be classified into two groups: the first group comprises traditional PMTs, some of which are capable of resolving up to 5 pe; the second group includes HPDs, MCPMTs, and SiPMs, for which up to 30 pe can be resolved. Interesting distributions are generated with m_0 set to about 2 for traditional PMTs; for other devices with superior resolution, m_0 is set in the region of about 5. With these parameters in mind, we evaluate (4.86) for $m_0 = 2$ and successively for $\delta_1 = 10, 15$, and 20 , with the results shown in Fig. 4.11. Figure 4.12 represents higher resolution devices with $m_0 = 5$, and $\delta_1 = 90$, typically. The spectra given in Figs 4.11 and 4.12 are instantly recognizable but they do not mimic experimentally determined MPHDS at the sub-photoelectron level.

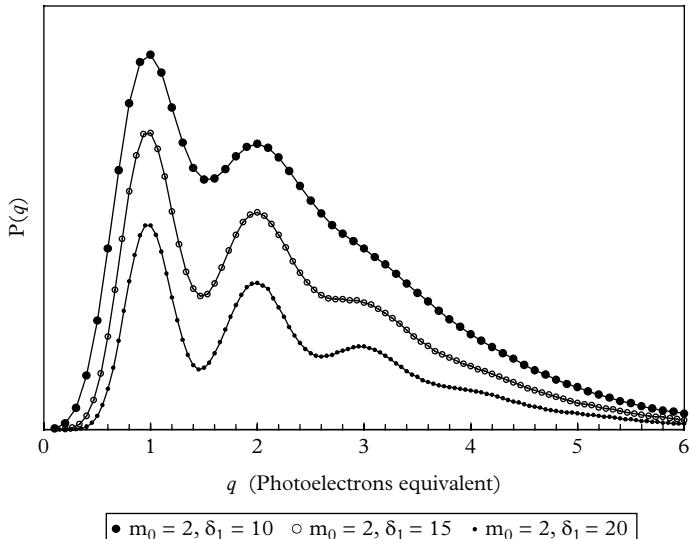


Fig. 4.11. Predicted MPHD based on (4.86) for a high d_1 gain PMT. The average number of photoelectrons per pulse is m_0 , and δ_1 is the assumed gain of the first dynode. Ordinates have been arbitrarily scaled for ease of comparison.

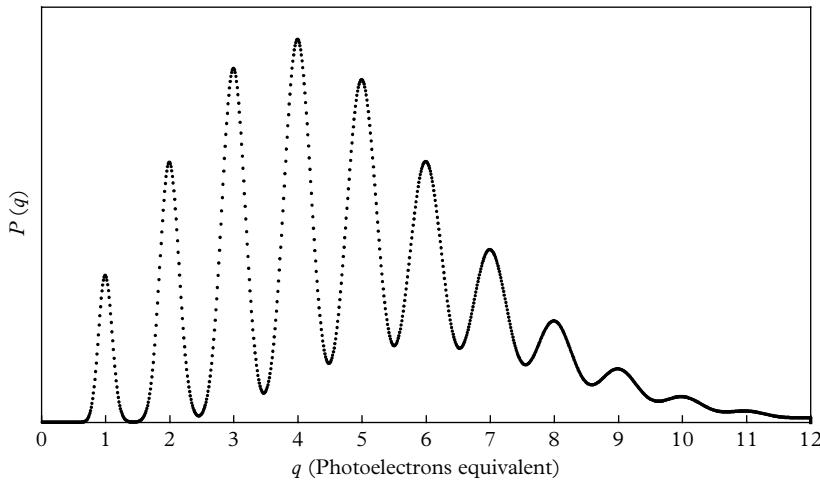


Fig. 4.12. The expected pdf for Group 2 devices derived from (4.86), with $m_0 = 5$, and $\delta_1 = 90$. The minima lift from the baseline because of the finite resolution of the SER, and the broadening of the peaks evident at high multiplicity.

4.15 MPHD for an empirical SER

The analysis that follows will reveal the internal structure of the output pulse height distribution envelope in terms of a set of $1, 2, \dots, N$ pe pulse height distributions, all weighted in accordance with the coordinates, $q(r)$, representing a given SER.

While (4.86) predicts the recognizable shape of the MPHD, the critical test lies in the ability of the theory to predict the actual pdf for a measured input distribution. In this case, we do not assume that Poisson statistics apply to the multiplier but an empirical SER is substituted instead thereby providing a set of 23 $q(s)$ coordinates for use with (4.85). In addition to this data, the mean number of photoelectrons, m_0 , per light pulse must be known. This can be determined in two ways, both of which should be used to give confidence in the assumed value for m_0 . The multi-photon pdf is measured at the same PMT gain as is the SER. In both cases, the mean anode current and the total number of counts per second are recorded. If the anode currents are I_e and I_m for single and multi-photon excitation respectively, with corresponding count rates of n_e and n_m , then $m_0 = I_m n_e / (I_e n_m)$ (see also 4.1.1). As a check, m_0 can be computed using the experimental data for the MPHD and for the SER. As they are measured at the same gain, m_0 is simply the ratio of the mean pulse heights of the two distributions. These methods result in $m_0 = 2.1$ for the distribution shown in Fig. 4.13. There is always a finite probability of recording no output, quantified by $q(0)$ and $P(0)$. However, experimentally determined SERs and MPHDs such as those shown in Figs 4.13 and 4.14 do not register zeroes.

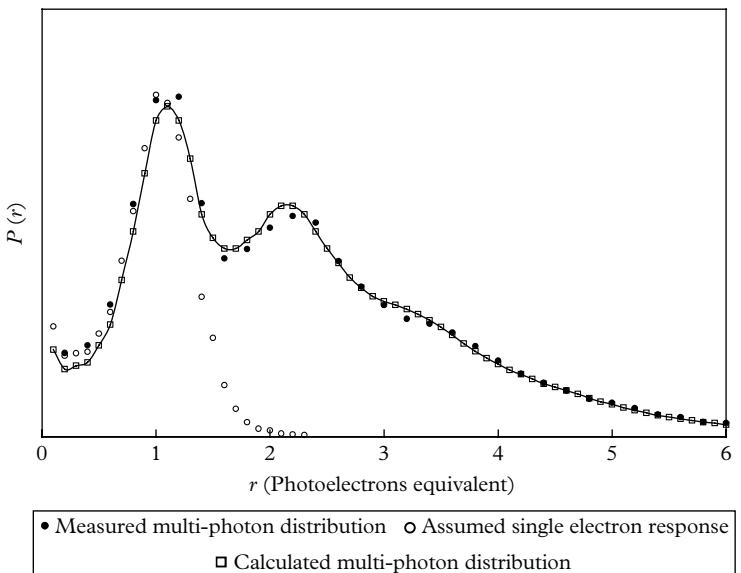


Fig. 4.13. Illustrating good agreement between the measured MPHD and (4.85). The assumed SER is the measured one for 23 $q(s)$ coordinates (open circles). Neither shifting nor normalization has been applied.

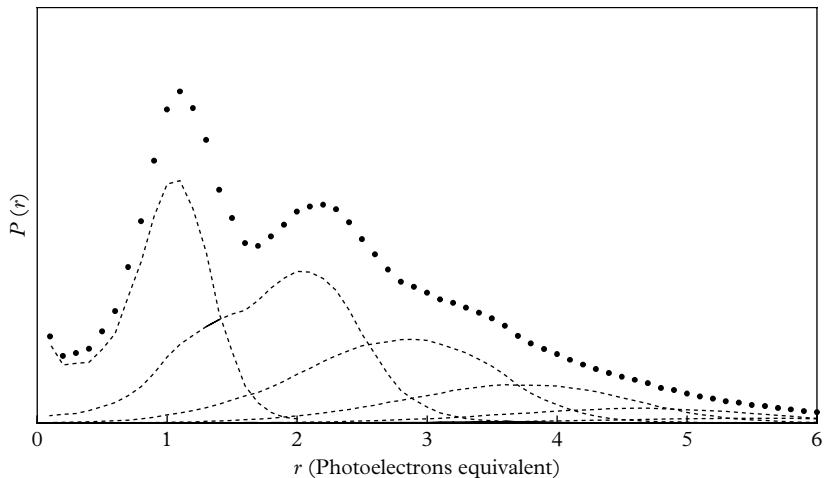


Fig. 4.14. The first six N -fold distributions from (4.85), weighted by Poisson coefficients $e^{-m_0} \times m_0^N / N!$. The sum of these curves follows the measured multi-photon distributions shown in this figure and in Fig. 4.13.

Wright A. G. (2005). The statistics of multi-photoelectron pulse height distributions. *Nucl. Instr. and Meth. in Phys. Res. A*, 579, 967–72. Reprinted from Elsevier Science ©2005.

4.16 Arrival time statistics

Determination of the precise time of occurrence of a light signal is a common requirement in both industrial and scientific applications of PMTs. For example, the emission of Cerenkov light, and scintillations, from certain organic materials are characterized by light signals of short duration—measured in nanoseconds. Repeated events, notionally the same in photon content, will always exhibit a spread in this number, and also in the time signature. In its most rudimentary form, timing with PMTs is based on generating a standard digital signal by triggering a pulse height discriminator. The question arises concerning where to set the discrimination level in terms of photoelectrons equivalent. It will be shown theoretically that, for best timing, discrimination should be set to detect the arrival of the first photoelectron. In practice, to accommodate a wide dynamic range in signal amplitudes, it is customary to use the zero crossing or constant fraction (CF) methods which are based on a multi-photoelectron threshold setting (see 8.5). The situation is different where digital signal processing is concerned, but this will not be explored. The presentation that follows is based on the analysis of Kelbert *et al.* (2006).

4.16.1 Exponential time signatures

We seek an expression for the emission-time distribution $P(t, Q | R)$ of the Q^{th} photoelectron, for Q ranging from 1 to R and subject to the boundary conditions: $F(0) = 0$; $F(\infty) = 1$. Two cases are considered: first, R is constant, which considerably simplifies computation. Second, of greater utility, R is assumed to be Poisson distributed; $F(t)$ describes cumulative emission of photoelectrons, while $F'(t)$ is the rate of photoelectron output—the time signature. In addition, $F(0) = 0$ implies that no photoelectrons are emitted for $t < 0$, while $F(\infty) = 1$ stipulates that the total number of emitted photoelectrons is finite, with $F(t)$ normalized to unity. The probability distribution, for the emission of the Q^{th} photoelectron out of R detected, at a time between t and $t + dt$, is best envisaged as a three-part process. Initially, R is fixed—that is, there is no statistical fluctuation. The three-part distribution is

$$P(t, Q | R) = p_a \times p_b \times p_c, \quad (4.87)$$

where:

p_a is the probability that $Q - 1$ photoelectrons arrive within $t = 0$ and t

p_b is the probability that the Q^{th} photoelectron arrives between t and $t + dt$

p_c is the probability that $(R - Q)$ photoelectrons (the remainder) arrive after $t + dt$

and

$$\begin{aligned} p_a &= C(R, Q-1) \times \left(F(t) - F(0) \right)^{Q-1} = C(R, Q-1) \times \left(F(t) \right)^{Q-1}, \\ p_b &= C(R-(Q-1), 1) \times F'(t) dt, \\ p_c &= \left(F(\infty) - F(t) \right)^{R-Q} = \left(1 - F(t) \right)^{R-Q}. \end{aligned}$$

Here $C(R, Q-1)$ is the number of combinations of R taken $Q-1$ at a time, and $C(R-(Q-1), 1)$ is the number of combinations of $R-(Q-1)$ taken one at a time, giving

4

$$P(t, Q | R) = \frac{R! \times \left(F(t) \right)^{Q-1}}{(Q-1)! \times (R-Q)!} F'(t) \times \left(1 - F(t) \right)^{R-Q} dt. \quad (4.88)$$

Consider an exponential decay for the optical signal, with a time constant $\tau = 1/\lambda$, as the excitation, $F(t)$, and its derivative $F'(t)$ as

$$\begin{aligned} F(t) &= 1 - \exp(-\lambda t), \\ F'(t) &= \lambda \exp(-\lambda t). \end{aligned} \quad (4.89)$$

The probability distribution for the arrival of the Q^{th} photoelectron out of R follows by substituting (4.89) into (4.88), giving

$$P(t, \lambda, Q | R) = \frac{R! \lambda \times (1 - e^{-\lambda t})^{Q-1}}{(Q-1)! \times (R-Q)!} \times e^{-\lambda(R-Q+1)t} dt. \quad (4.90)$$

We can derive the probability expression for R , Poisson distributed, with a mean R_m of the form $p(n) = (R_m)^n \times \exp(-R_m)/n!$ for inclusion in (4.90); thus,

$$\begin{aligned} P(t, \lambda, Q | R_m) &= \sum_{R=Q}^{\infty} p(n) \times P(t, \lambda, Q | R) \\ &= \sum_{R=Q}^{\infty} \frac{(R_m)^R}{R!} \times \frac{e^{-R_m} R! \times \lambda (1 - e^{-\lambda t})^{Q-1}}{(Q-1)! \times (R-Q)!} \times e^{-\lambda(R-Q+1)t} dt \\ &= e^{-R_m} \lambda \times \frac{(1 - e^{-\lambda t})^{Q-1} \times e^{-\lambda t} \times (R_m)^Q}{(Q-1)!} \\ &\quad \times \sum_{R=Q}^{\infty} \frac{(R_m)^{R-Q} \times e^{-\lambda(R-Q)t}}{(R-Q)!} dt \\ &= e^{-R_m} \lambda \times \frac{(1 - e^{-\lambda t})^{Q-1} \times e^{-\lambda t} \times (R_m)^Q}{(Q-1)!} \times \sum_{R=Q}^{\infty} \frac{(R_m)^k \times e^{-\lambda kt}}{k!} dt \\ &= (R_m)^Q \lambda \times \frac{(1 - e^{-\lambda t})^{Q-1} \times e^{-\lambda t}}{(Q-1)!} \times \exp(-R_m \times (1 - e^{-\lambda t})) dt. \end{aligned} \quad (4.91)$$

The timing distributions for (4.90) and (4.91) are given in Fig. 4.15 for $R = 10$, where the time constant of a NaI(Tl) scintillator is assumed.

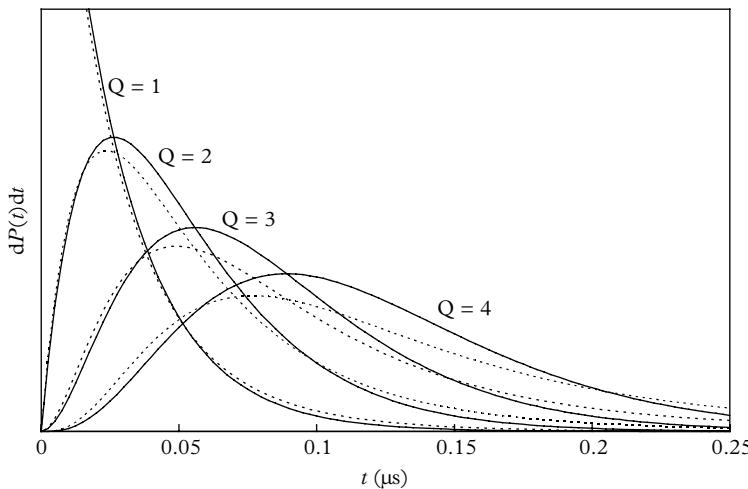


Fig. 4.15. Timing distributions for the arrival of the first, second, third, and fourth photoelectrons in the decay of NaI(Tl) with $\tau = 250$ ns. Solid curves are derived from (4.90) with $R = 10$, and the dashed curves from (4.91) with $R_m = 10$ pe, Poisson distributed.

Wright A. G. (2005). The statistics of multi-photoelectron pulse height distributions. *Nucl. Instr. and Meth. in Phys. Res. A*, 579, 967–72. Reprinted from Elsevier Science ©2005.

4.16.2 First two moments for R fixed

We make use of the fact that the time intervals between emissions are independent. The time $t(Q)$ to emit Q photoelectrons can be represented as a sum of independent random variables

$$t(Q) = t_1 + t_2 + \dots + t_Q = \sum_{k=1}^Q t_k, \quad (4.92)$$

where t_k is the time between the emission of the $(k-1)^{\text{th}}$ and the k^{th} photoelectrons. Every t_k in (4.92) has an exponential distribution with its own parameter $\lambda(R-k+1)$, with the pdf for t_k taking the form

$$p_k(t) = \lambda(R-k+1) \times \exp(-\lambda(R-k+1)t).$$

We avoid cumbersome combinatorial arguments in what follows by utilizing the statistical independence of t_k . The mean and variance are obtained as follows:

$$E[t_k] = \int_0^\infty p_k(t) t dt = \frac{1}{\lambda(R-k+1)},$$

$$\text{var}(t_k) = \int_0^\infty p_k(t)(t - E[t])^2 dt = \frac{1}{\lambda^2(R - k + 1)^2},$$

leading to

$$E[t(Q)] = \sum_{k=1}^Q t_k = \sum_{k=1}^Q \frac{1}{\lambda(R - k + 1)} \quad (4.93)$$

4

and

$$\begin{aligned} \text{var}(t(Q)) &= \text{var}(t_1) + \text{var}(t_2) + \dots + \text{var}(t_k) \\ &= \sum_{k=1}^Q \frac{1}{\lambda^2(R - k + 1)^2}. \end{aligned} \quad (4.94)$$

The mean arrival times $E[t(Q)]$ of the first and second photons, for example, from an emission of a total of 10 photons, are 25 ns and 52.8 ns, respectively. This follows from (4.93) and is based on an assumed time constant of 250 ns. The corresponding arrival time distribution is given in Fig. 4.15. Note that we have Q terms in (4.94) and, if $Q \ll R$, then all terms have approximately the same magnitude of $1/(\lambda^2 R^2)$, leading to

$$\text{var}(t(Q)) \approx Q/(\lambda R)^2$$

4.16.3 First two moments for R randomized

The mean emission time of the Q^{th} photoelectron, $E[t(Q, R_m)]$, and its variance is given by

$$Et[Q|R_m] = \sum_{R=Q}^{\infty} p(R) \times \sum_{k=1}^Q \frac{1}{\lambda(R - k + 1)} \quad (4.95)$$

and

$$\begin{aligned} \text{var}(Q|R_m) &= \sum_{R=Q}^{\infty} p(R) \times E[R^2] - (E[t(Q)])^2 \\ &= \sum_{R=Q}^{\infty} p(R) \times \left\{ \sum_{k=1}^Q \frac{1}{\lambda^2(R - k + 1)^2} + \left[\sum_{k=1}^Q \frac{1}{\lambda(R - k + 1)} \right]^2 \right\} \\ &\quad - \left[\sum_{R=Q}^{\infty} p(R) \times \sum_{k=1}^Q \frac{1}{\lambda(R - k + 1)} \right]^2. \end{aligned} \quad (4.96)$$

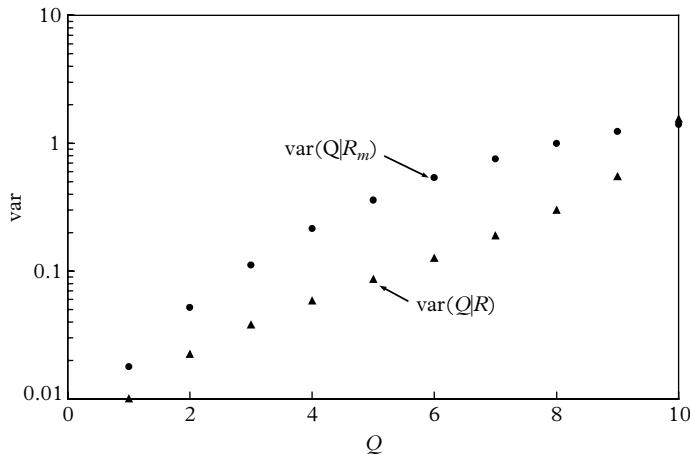


Fig. 4.16. Variance of emission time, for $R = 10$ (fixed) and for R Poisson distributed with a mean of $R_m = 10$. The entries are computed from equations (4.94) and (4.96) respectively, with $\lambda = 1$.

Arrival time statistics. Kelbert, M., Sazonov, I. and Wright, A. G. (2006). Exact expression for the variance of the photon emission process in scintillation counters. *Nucl. Instr. and Meth. in Phys. Res. A*, 564, 185–9. Reprinted from Elsevier Science ©2006.

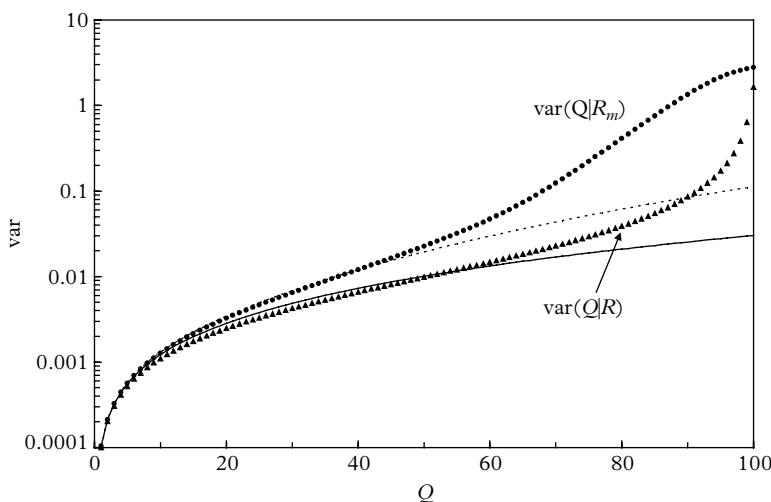


Fig. 4.17. The variance for $R = 100$, fixed, (4.94), and variable, (4.96), solid dots. The continuous line represents $Q/(\lambda^2 R_m^2)[1 + 2(Q + 1)/R_m]$, while the dashed curve refers to the four-term expansion, (4.97).

The formulations in (4.94) to (4.96) are exact but unwieldy in computation, with sample results shown in Figs 4.16 and 4.17 for R fixed and for R randomized. Note that for R fixed, Q is always less than or equal to R but, for R randomized, $\text{var}(Q, R_m)$ is also defined for $Q > R_m$. However, the points for $Q > R_m$ have not been shown in the figures.

A method for devising a closed formula representing (4.96) is given by Kelbert *et al.* (2006) as follows:

$$\begin{aligned} \text{var}(Q | R_m) = & \frac{Q\tau^2}{R_m^2} \left(1 + \frac{2(Q+1)}{R_m} + \frac{(Q+1)(6Q+11)}{R_m^2} \right. \\ & \left. + \frac{(Q+1)(Q+2)(2Q+5)}{R_m^3} \right). \end{aligned} \quad (4.97)$$

Post and Shiff (1950) presented a formula with the same form of (4.97) but limited to the first two terms within the square brackets.

Best timing is achieved by detecting the arrival of the first photon and ignoring the contribution to jitter by the PMT. This is assessed in 8.7.2.

4.17 Determination of afterpulse rates

An ideal PMT produces a single charge pulse at the anode for each photoelectron collected. The pulse extends over a few nanoseconds in fast PMTs and for those classified as slow, the pulse duration is of the order of 20 ns. However, all PMTs suffer to some degree from afterpulses. As this term suggests, there is sometimes a satellite pulse, allied in time, with the signal pulse. A delayed coincidence measurement can reveal the time distribution and the afterpulse rate; an autocorrelator may also be used for this purpose. Afterpulse rates can be determined more readily with a basic laboratory scaler if, as in photon counting, interest lies only in determining the afterpulse rate and not its spectrum. Afterpulse probabilities range from 0.01 % to about 5 % per photoelectron, depending on the type of PMT and its operating conditions: they occur predominantly between 20 ns and 10 μ s following a signal pulse. Pulse heights range between a fraction of a single-photoelectron signal to about 10 photoelectrons equivalent. A detailed account of afterpulses is reserved for 11.5, while the present objective is to investigate statistical aspects only.

A histogram in 4.5.4 comprising 10,000 samples of a random light source serves to illustrate the various statistical computations. The variance and the mean are identical for a Poisson source of light viewed by a perfect PMT. This equivalence does not apply to PMTs exhibiting afterpulses. In this case, the variance exceeds the mean and the afterpulse rate follows by determining the difference between these two parameters.

4.17.1 Afterpulses: Method I

The afterpulse rate, p_a , is the probability that a single-photoelectron event produces a satellite pulse of sufficient amplitude to re-trigger the counting electronics. The theory is developed on the following basis: a signal pulse produces either one afterpulse with probability p_a , or, no afterpulse, with probability $(1 - p_a)$. If the light source is sufficiently weak, the probability of more than one afterpulse per event may be taken as zero. This is another example of a cascade process in which A refers to the light source producing on average μ photoelectrons per sample time; B acts on these photoelectrons, generating an occasional afterpulse subject to binomial statistics. The generating functions are

$$G_A(u) = e^{-\mu} \times \exp(\mu u),$$

$$G_B(u) = (1 - p_a)u + p_a u^2,$$

$$G_{AB}(u) = G_A[G_B(u)],$$

$$G_{AB}(u) = e^{-\mu} \times \exp\left(\mu(1 - p_a)u + \mu p_a u^2\right),$$

$$G_{AB}(1) = 1,$$

$$G_{AB}'(u) = G_{AB}(u) \times \left(\mu(1 - p_a) + 2\mu p_a u\right),$$

$$G_{AB}'(1) = \mu(1 + p_a),$$

where the latter is the measured mean, compared with ideal performance where $p_a = 0$;

$$G_{AB}''(u) = G_{AB}'(u) \times \left(\mu(1 - p_a) + 2\mu p_a u\right) + G_{AB}(u)2\mu p_a,$$

$$G_{AB}''(1) = \mu^2(1 + p_a)^2 + 2\mu p_a,$$

$$\text{var}_{AB} = \mu^2(1 + p_a)^2 + 2\mu p_a + \mu(1 + p_a) - \mu^2(1 + p_a)^2.$$

The first and last terms on the right cancel and

$$\text{var}_{AB} = \mu(1 + 3p_a),$$

compared with a variance of μ when $p_a = 0$.

We can determine the afterpulse rate p_a by computing var_{AB} and the mean, $G_{AB}'(1)$, for substitution into (4.98):

$$\frac{\text{var}_{AB}}{\text{m}_{AB}} = \frac{1 + 3p_a}{1 + p_a}. \quad (4.98)$$

Rearranging (4.98) leads to an expression for the afterpulse rate in terms of measured parameters, var_{AB} and $G_{AB}'(1)$:

$$\begin{aligned} p_a &= \frac{\left(\frac{\text{var}_{AB}}{\text{m}_{AB}} - 1\right)}{\left(3 - \frac{\text{var}_{AB}}{\text{m}_{AB}}\right)} \\ &\approx \frac{1}{2} \left(\frac{\text{var}_{AB}}{\text{m}_{AB}} - 1\right). \end{aligned} \quad (4.99)$$

It is assumed in the last step that we are concerned with a usable PMT, in which case $\text{var}_{AB}/\text{m}_{AB}$ is close to unity; this expression is also given by Oliver (1977). It is obvious from (4.99) that low afterpulse rates are difficult to measure accurately since p_a is determined by the difference of two nearly equal quantities, making $\text{var}_{AB}/\text{m}_{AB}$ close to unity. Any shift from stable performance has a disproportionate effect on computation, where afterpulse rates are low. Note that no contribution has been included from the creation of afterpulses by the afterpulses themselves. While this does occur, the contribution to the afterpulse rate is negligible if p_a is less than ~ 0.1 , and arguably within the upper performance expectation for a ‘quality’ PMT.

4.17.2 Afterpulses: Method II

This method, also based on a single-photon source, takes advantage of the observation that afterpulses are largely confined within 10 μs of the initiating pulse. Meeting the condition given in 4.1.1 for producing a single-photoelectron source, by ensuring that $p(2)/p(1) \ll \mu/2$, implies that all dual counts arise predominantly from afterpulses. If $P(n)$ represents the probability of recording n events in the presence of afterpulses, and $p(n)$ is the Poisson probability that applies in the absence of afterpulses, then

$$P(0) = p(0) = \exp(-\mu),$$

$$P(1) = p(1)(1 - p_a),$$

$$P(2) = p(1)p_a + p(2)(1 - p_a)^2,$$

$$\frac{P(2)}{P(1)} = \frac{p(1)p_a + p(2)(1 - p_a)^2}{p(1)(1 - p_a)}$$

$$= \frac{p_a}{1 - p_a} + \frac{\mu}{2}(1 - p_a).$$

If the light intensity is adjusted such that $\mu \ll p_a$, we can ignore the second term, putting $P(2)/P(1) = \alpha$, which leads to a simple formula for the determination of p_a :

$$p_a = \alpha/(1 + \alpha) \approx P(2)/P(1). \quad (4.100)$$

In (4.100) all paired events are assumed to be afterpulses. Generally, if the afterpulse rate lies between 1 and 10 %, μ must be less than 0.002 for the assumption to hold. To measure PMTs with exceptionally low afterpulse rates, such as the ET 9863, μ needs to be set at 0.0002 to achieve the stated accuracy. The measurement precision depends upon the fluctuation in $P(2)$ since $P(1)$ is subject to a smaller error. To determine the afterpulse rate to within $\pm 10\%$ requires only about 100 $P(2)$ events.

An example of an afterpulse determination based on (4.99) or (4.100) is given in Chapter 11 in Fig. 11.10. The advantages of Method II for determining an afterpulse rate, as compared with Method I, are as follows:

- It is insensitive to drift in the strength of the light source or gain shift.
- It is relatively immune to interference from the occasional burst of pulses.
- The data computation for it is simpler than that for Method I, as there is no need to calculate the variance.

4.18 Scaling and dead time

Counting digital pulses is the basis of an important class of instrumentation: familiar examples of instruments that do this are photon counters, luminometers, and nuclear radiation probes. Software may be included for manipulating the data to present it in the form the user requires: possibly graphical or perhaps expressed as a concentration of some biochemical, or the activity of a particular radioisotope. The primary information is a count rate, whatever transformations are subsequently performed. It is wise to question all imbedded data manipulations, for they may be neither appropriate nor statistically justifiable. In particular, be wary of a commercial detector head that includes a divide-by-p scaler. Manufacturers describe this option as ‘prescaled counts’ (typically achieving a reduction by a factor of 2, 4, or occasionally 10), but the circumstances under which scaled

counts may offer an advantage are not always clear. A software procedure is sometimes followed that ignores the highest and the lowest of a set of repeated counts, on the basis that a high reading may have been induced by electrical interference: this approach can be justified in certain circumstances, for example, where PMT background counts are known to be non-statistical. The application of this technique is fraught with possible statistical difficulties, and users need to be wary in its adoption.

4.18.1 Interval distribution

A formulation that describes the distribution in length of the time intervals between successive events is derived below. Consider a constant mean rate of a events per unit time, and assume Poisson statistics. The probability of recording n counts within unit time is

$$P_n = \frac{(at)^n}{n!} \times e^{-at} \quad (4.101)$$

The probability that there will be no events in a time interval t , during which we expect at events on average, is

$$P_0 = \frac{(at)^0}{0!} \times e^{-at} = e^{-at}. \quad (4.102)$$

The probability of recording an event in the time interval dt is simply adt . The combined probability that there will be no events during the time interval t but one event between time t and $t + dt$ is from (4.102):

$$dP_t = e^{-at} a dt. \quad (4.103)$$

It is interesting that dP_t is a continuous function: that is, all time intervals are permitted. In contrast, in (4.101), only integers are allowed—half counts, for example, do not exist. What is even more interesting is that the most likely time interval between events is zero, and not $1/a$ as the unwary may intuitively assume. Figure 4.18 illustrates (4.103).

To acquire a graphical representation of the randomness of events, from a light source of constant intensity, time intervals are selected in accordance with (4.103) using random numbers. Figure 4.19 represents the theoretical arrival times from a constant intensity but random photon source. The mean rate is one per second, measured over a period of 100 seconds. The bunching effect predicted by (4.103) is evident.

4.18.2 The p-fold time interval

The reliance on scaling was well known to the nuclear physics community of the 1950s. A digital scaling circuit produces one output pulse for every p input pulses. Cascading k flip-flop circuits achieves division of the count rate by 2^k , and decade

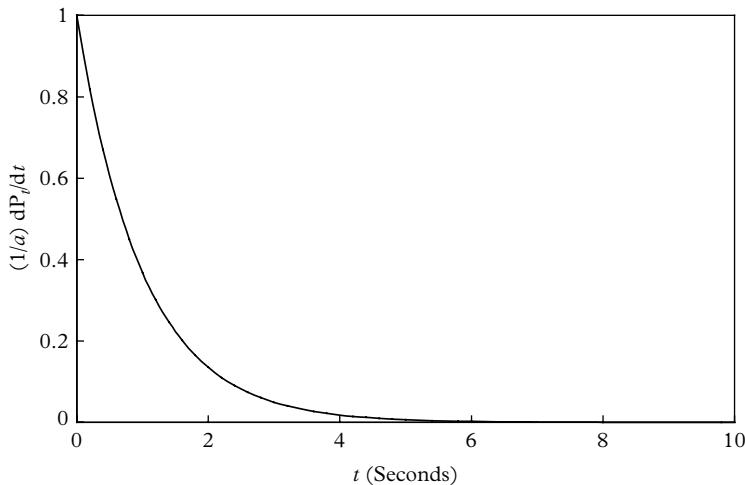


Fig. 4.18. Time interval distribution for a random source of constant intensity given by $(1/a) (dP_t/dt) = e^{-at}$, where a has been taken as 1 count/s. The most probable time interval between events is 0 s, and the average time interval is 1 s. The area under the curve is unity, as may be verified by integration.

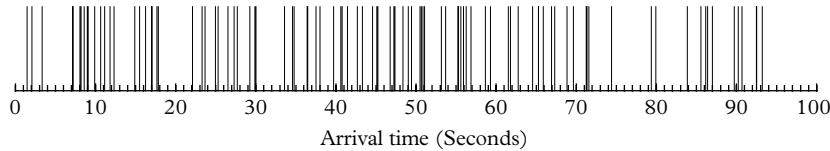


Fig. 4.19. A typical arrival time pattern for a source of constant mean intensity of 1 count/s. The tendency for events to bunch, as predicted in Fig. 4.18, is evident. Every repeated time sequence is different, and a further 100 second sequences will all have unique arrival time signatures. The common factor is the set of time intervals, which, if binned appropriately, will follow (4.103). The pattern shown is representative of any random count rate, following rescaling the abscissa.

scalers can be realized by making appropriate interconnections between the flip-flops. Historically, the need for scaling was to reduce the count rate, making it compatible with the reaction time of electro-mechanical and glow tube registers, which at that time were the only means for storing counts. Contemporary timer counters have advanced considerably and general laboratory instruments now offer typically 100 MHz counting capability and scalers that are more specialized, operate at gigahertz rates.

A general formula for the p-fold distribution, which is the distribution that results from dividing the counts by any whole number p , is easily derived and shown in Fig. 4.20. The length of an interval between output pulses is termed the

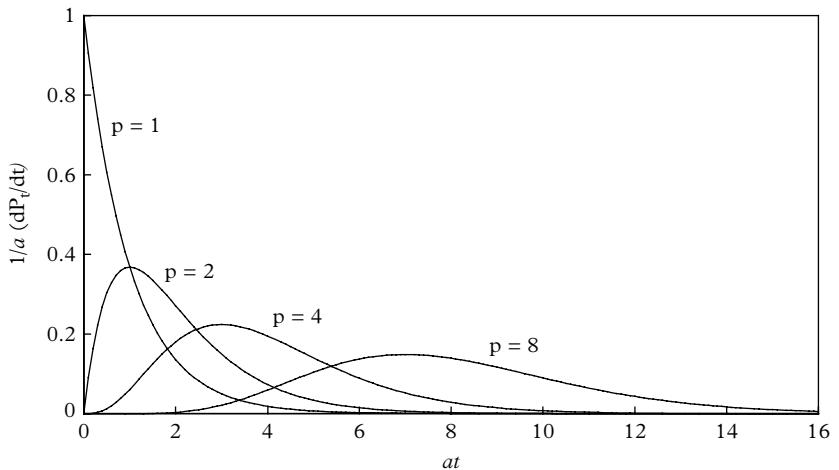


Fig. 4.20. Time interval distributions generated by scaling, based on (4.104). The abscissa is the average number of events at in a time interval t . The scaling factor p may be any whole number.

p -fold interval. The intervals between successive input pulses contain zero events but the intervals between successive p -fold events contain $(p-1)$ events. The Poisson distribution that predicts the probability that a p -fold interval of duration t contains exactly $p-1$ events is

$$P_{p-1}(t) = \frac{(at)^{p-1}}{(p-1)!} \times e^{-at}.$$

The probability of one event occurring in an additional time dt is simply $P_1(dt) = adt$.

The probability of $p-1$ events in t , and the p^{th} event between t and $t+dt$, is $dP_t = P_{p-1}(t) \times P_1(dt)$,

$$P_t = \frac{a^p t^{p-1}}{(p-1)!} \times e^{-at} dt, \quad (4.104)$$

which is the generalized p -fold distribution. These distributions are plotted in Fig. 4.20, where for convenience a is taken as unity. Note that (4.104) has the same form as (4.20), the factorial function. We can illustrate the result of scaling graphically, by applying it to the time sequence of Fig. 4.19. The second row from the bottom of Fig. 4.21 is constructed by showing every other event in the bottom row, thus mimicking division by 2. Similarly, the third row is generated from every other event in the second row to yield the distribution for $p = 4$, and so forth.

There are several points to note in connection with scaling:

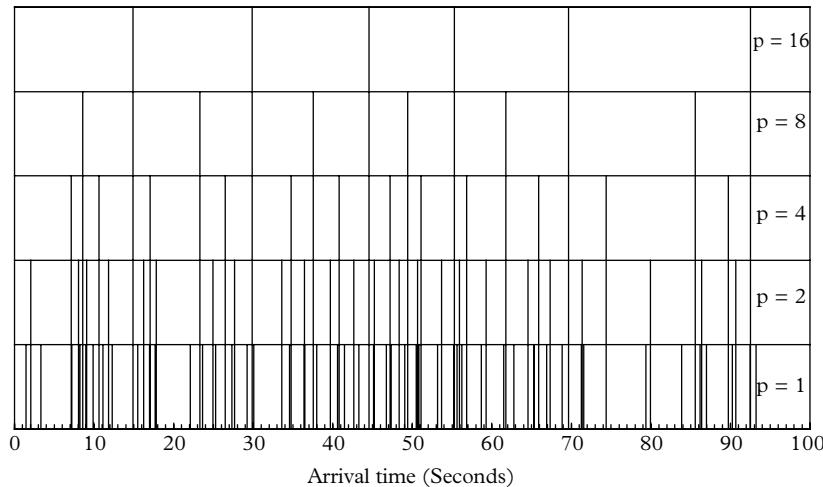


Fig. 4.21. Illustrating how scaling removes short time intervals from a sequence of random events, hence easing the demands on the required response time of any associated counting equipment. The regularizing action produced by scaling is evident as p is increased.

4

- The duration of the most probable p -fold interval (corresponding to the peaks) is $(p - 1)$.
- The average p -fold interval is p/a , as expected.
- The area under each curve between $t = 0$ and $t = t$ is the probability that a p -fold interval will be shorter than at .
- There is a significant reduction in the occurrence of short intervals; this is produced by even small scaling factors.
- There is a strong regularizing effect for the larger scaling values—the interval distribution becomes noticeably symmetrical with increasing p (e.g. $p = 8$).

Scaling results in diminished precision, with obvious consequences in signal-minus-background calculations. This is particularly serious whenever signal and background counts are similar in magnitude. Precision is further compromised if count rates are also low. Mean counts and variance should be deduced from unscaled data wherever possible. Some manufacturers of commercial instrumentation use scaling to promote customer confidence in the repeatability of measurements. Operators who are not versed in statistics may interpret display flicker as a sign of faulty instrumentation. Flicker is readily ameliorated by scaling.

4.19 Dead time

All instruments used for counting digital pulses have a characteristic recovery time during which the counter is disabled from counting—it is said to be ‘busy’, ‘paralysed’, or ‘dead’. After recording one pulse, a counter is unresponsive to successive pulses until a time interval equal to or greater than its resolving time has elapsed. This element of time is termed dead time, τ , and it aids understanding of later sections to consider a counter as an instrument that operates on the intervals between events rather than on the events themselves. The interval distribution (4.104) gives the theoretical allocation of time intervals between photons emitted by a steady light source obeying Poisson statistics and it also applies to photo-emission (see 4.10.2). The particular property to note is that short intervals between successive events are more frequent than longer intervals, with the most likely time interval of zero. This gives rise to ‘bunching’, evident in Fig. 4.19, in which sets of closely spaced, yet random counts, appear. If the interval between two events is shorter than τ , the second event will not be recorded, and clearly there is a link between bunching of events and loss of counts.

The treatment of dead time in the sections that follow focusses on photon counting, but it is sufficiently general for application to other types of detector systems, such as scintillation counters. The subject of dead time has a long history dating back to the time when physicists first needed to count electronic pulses. Its importance grew in line with the rapid growth of the nuclear industry in the 1950s. The classic textbook *The Atomic Nucleus* by Evans (1955) still serves as an excellent primer for those seeking an erudite introduction to the subject. Even those readers familiar with dead time will appreciate Evan’s rigorous treatment and his references to early work on the subject.

4.19.1 Sources of dead time

Dead time manifests itself in two ways: it is present in the PMT itself and in any associated counting instrumentation. According to the interval distribution, there are always some unresolved pairs of events for which pairs of output pulses will appear as a single large amplitude one, or more frequently, as a broad one showing two peaks. This overlap mechanism is usually referred to by the descriptive term ‘pulse pile up’. It is clear in the representations of Fig. 4.22 that PMT pulses separated by less than about 2.5 ns will appear coincident where a threshold detector is used. Counting pulses always involves setting the threshold of a discriminator (shown by the dashed line) such that only those pulses exceeding this level are counted. The pulses in the first pair are sufficiently separated to enable the discriminator to fire twice. In the second pair of pulses, the second pulse barely fires the discriminator. The time separation of the third pair is insufficient to resolve the two pulses, and a single count is registered. The time separation of the

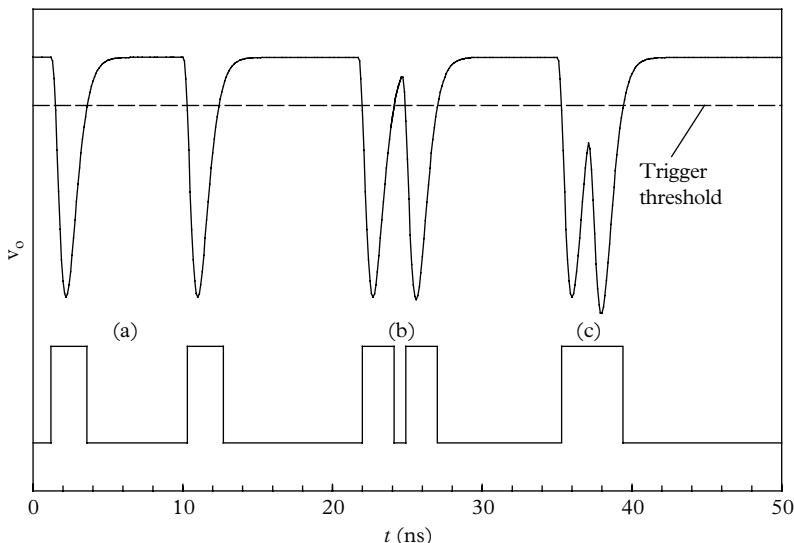


Fig. 4.22. Showing how pulse overlap of anode pulses causes loss of counts. The lower trace represents the output from an ideal discriminator. The first pair of pulses, indicated by (a), is adequately separated in time, and both are registered if the dead time of the discriminator is less than ~ 7 ns. Counting both pulses is unlikely with the second pair, indicated by (b), because of the dead time of the discriminator. Only a single pulse is generated from the third pair, (c).

anode pulses (peak to peak) has to be greater than ~ 2.5 ns in order to register both pulses, in this example. Equivalently, the dead time is of the order of $2 \times t_{\text{fwhm}}$.

The essential characteristic of a discriminator, whether it is a threshold or CF type, is the generation of output in the form of a logic pulse of fixed height and width. Dead time is set by an internal monostable, operating in the range of 1 to 30 ns. Electronic circuitry must have recovered prior to the arrival of the next input. In other words, the discriminator has its own dead time regardless of any PMT pulse overlap losses. The third element in all counting systems is the means to record the counts produced by a discriminator: such instruments are referred to as timer counters or scalers. These instruments have similar limitations on count rate as do discriminators, and if the gap between logic pulses is too small, such as that depicted in Fig. 4.22(c), the second signal fails to record.

Although the effects of fluctuations in pulse height and shape of PMT signals have been ignored, this treatment is sufficient to indicate what might be expected from the combined performance of a PMT and associated electronics. The specification for a commercial scaler may be quoted as ‘100 MHz counting capability’. Such information is useless as a specification unless test conditions are clearly stated: pulses should be random in time and of stated width. The user,

given a high-quality pseudorandom pulse generator, may readily check manufacturer's information.

The description of dead time losses depicted in Fig. 4.22 refers to the traditional way of counting photoelectrons. With digital signal processing it is possible to disentangle closely spaced pulses, such as those shown in Fig. 4.22(b) and (c), permitting correction of higher count rates than previously mentioned. This is a specialized subject on which further information concerning digital signal processing applied to X-ray and gamma-ray detectors may be found in one of the earlier papers on the subject (Stein *et al.* 1996).

4.19.2 Counting losses attributed to dead time

It is clear from the discussion in the previous section that finding an exact correction for dead time presents difficulties. However, exact formulations can be derived for two ideal types of discriminator and these can be adapted for practical purposes. Discriminators are referred to as 'paralysable' (also known as Type I) and 'nonparalysable' (or Type II). The essential difference in performance relates to the way in which each responds to pulses arriving during their dead time. The paralysable type is unable to generate a second output pulse unless there is a time interval of at least τ between pulses. During the response time, τ , to an initial event, the recovery of the counter is extended by a time equal to τ , by all events that occur before full recovery has taken place. This has the effect of extending the busy time, as illustrated in Fig. 4.23. Type II counters are unaffected by events that occur during the recovery time, τ , and if the observed count rate is n , the fraction of the unit run time during which the counter is dead is $n\tau$; the fraction of live time available to the counter is accordingly $(1 - n\tau)$.

4.19.3 Type I, paralysable counters

For a set of intervals N , the number detected between time t_1 and t_2 is given by (4.103) as

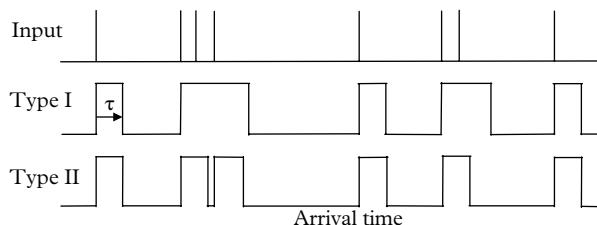


Fig. 4.23. An illustration of the different behaviours of Type I and Type II counters, each with finite resolving time τ . Infinitesimally narrow input pulses have been assumed.

$$\begin{aligned} n &= N \int_{t_1}^{t_2} e^{-Nt} dt \\ &= N \left(\exp(-Nt_1) - \exp(-Nt_2) \right). \end{aligned}$$

The number of intervals greater than $t_1 = \tau$ follows by letting $t_2 \rightarrow \infty$ in this expression:

$$n = N \times e^{-N\tau}$$

or

$$n\tau = N\tau \times e^{-N\tau}. \quad (4.105)$$

It can be shown by differentiation that (4.105) reaches a maximum when $N\tau = 1$ and that the lack of monotonicity has serious practical implications. At present, we seek an expression to correct a measured count rate, n , to provide an estimate of N , the true count rate. In (4.105) note that n is a function of N , $n(N)$, but we want the converse, $N(n)$. Rearranging (4.105) and expanding the exponential, taking the initial two terms only, leads to the first approximation:

$$N\tau = n\tau \times e^{N\tau} = n\tau(1 + N\tau + \dots) = n\tau/(1 - n\tau). \quad (4.106)$$

This will be recognized as one of the standard forms for dead time correction, albeit approximate. Note that at low count rates $N \approx n$ and (4.106) can also be expressed as $N\tau \approx n\tau(1 + n\tau)$. Equation (4.106) can be improved upon by including the third term of the expansion; thus,

$$N\tau = n\tau(1 + N\tau + (N\tau)^2/2! + \dots).$$

This is a quadratic in N , with a solution of the form

$$N\tau = \frac{1 - (1 + y)^{1/2}}{n\tau/(1 - n\tau)}$$

with

$$y = -2(n\tau)^2/(1 - n\tau)^2.$$

Expanding $(1 + y)^{1/2} = 1 + \frac{1}{2}y + \frac{1}{2}(\frac{1}{2} - 1)y^2/(2!) + \dots$ leads to

$$N\tau = \frac{n\tau}{(1 - n\tau)} + \frac{1}{2} \left(\frac{n\tau}{1 - n\tau} \right)^3 + \dots \quad (4.107)$$

The author acknowledges Gennadi Feltman of Soreq NRC, Israel, for the development of (4.107).

4.19.4 Type II, non-paralysable counters

As the name implies, the non-paralysable counter is unaffected by events that occur during its recovery time, τ . If the observed counts are n , then the fraction of running time that the counter is alive is $(1 - n\tau)$, and true counts are given by

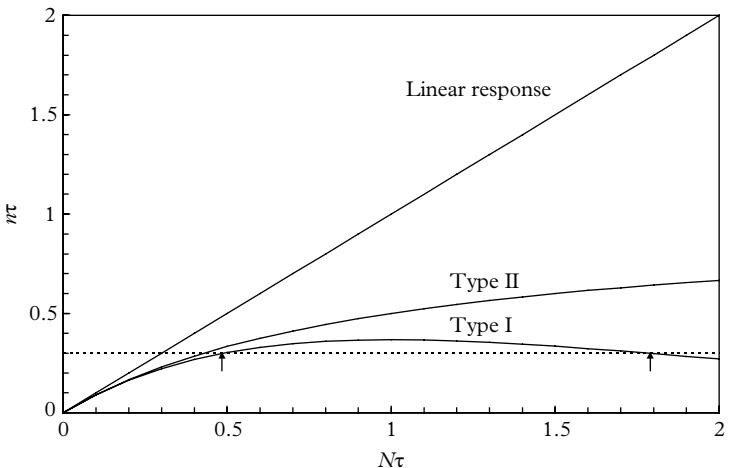


Fig. 4.24. Illustrating formulae (4.105) and (4.108) together with a counter of zero dead time. Type I counters are limited to a maximum count rate of $n = (et)^{-1}$, which occurs at $N\tau = 1$. There is always ambiguity with Type I counters: every reading of $n\tau$ counts corresponds to two values for N , with an example indicated by the arrows.

$$N = n/(1 - n\tau). \quad (4.108)$$

Equations (4.105) and (4.108) are plotted in Fig. 4.24, together with the response for a counter with zero dead time. Note that (4.108) can be inverted to express n as a function of N ; thus,

$$n = N/(1 + N\tau).$$

There are several points to note:

- At low count rates, $N\tau \ll 0.1$, all formulations converge.
- A non-paralysable counter approaches $n\tau = 1$ asymptotically.
- The paralysable counter attains a maximum count rate at $N\tau = 1$ but the non-monotonic response implies ambiguity, as illustrated in Fig. 4.24.

In Fig. 4.23, photo detections occur at the times shown by the thin vertical lines. Type I counters respond only to intervals greater than τ , while Type II instrumentation is dead for a time of τ after each input, but it can respond even if the interval between successive inputs is less than τ . Note in particular the response of the two types of counter to the triplet input. There are eight input counts in this example, of which five are recorded by a paralysable counter, and six by a non-paralysable one. Furthermore, it should be noted that the output ‘logic’ pulses from a Type II are consistent in width, τ , while those from a Type I are variable,

but never less than τ . The approach so far adopted takes no account of the inevitable dead time of the counting equipment itself. Even if a fast, state-of-the-art, 32-bit timer counter is used for this purpose, one needs to know the details concerning its pulse-pair resolution, as noted in 4.19.1. For example, counting the closely spaced doublet in the Type II mode of operation shown in Fig. 4.23 requires a counter of sufficient bandwidth to react to the narrow separation of the two constituent pulses.

It should be clear that dead time correction is necessarily a complicated subject, primarily because there are multiple contributions from various sources:

4

- pulse overlapping at the anode because of the finite width of PMT output pulses
- pulse height discriminator processing time (an essential item in any photon-counting system)
- resolving time of scaler timers or other means for counting digital pulses

The author's experience in the design of discriminators is that the actual performance falls between the two types and lock-up threatens at sufficiently high count rates in all counters. Manufacturers appreciate the importance of distinguishing between a genuine low count rate and an extremely high one masquerading as a low one. It would appear that a trigger circuit, activated whenever the anode output current exceeds a set level, would suffice to remove the ambiguity, yet the author is unaware of any published circuits that serve this purpose.

4.19.5 Dead time and variance

If N is the true rate of a random process, derived from measuring n over a time T , then the average number of observed events, nT , is smaller than NT by an amount which depends on the resolving time τ of the counter, the average rate n , and the counter classification (Type I or Type II). Variance, however, attributed for the true number of counts, is clearly not nT but is smaller than this number because of the loss of small time intervals. Kosten (1943), Feller (1948), and Evans (1955) give expressions for the variance of a Poisson distribution modified by dead time. Evans gives the variance for Type II counters as

$$\sigma_n^2 \approx NT/(1 + N\tau)^3$$

and for Type I we have

$$\sigma_n^2 = NT \times (1 - 2N\tau + N\tau^2/T)$$

There is little evidence in current literature for the adoption of these formulations and there is the dilemma as to which formulation applies—Type I or Type II?

4.19.6 Determination of dead time

The review paper by Müller (1973) is worth mastering for those wishing to take the subject further. The article refers to the ways for determining this parameter, such as the two-random-source method; a method using one random source plus a fixed frequency oscillator; and the decaying-source method. Two of these methods are described in detail by Knoll (1989) at a level that is comprehensible. A major difficulty highlighted by Müller is that absolute count rates are in general unknown but, fortunately, where PMTs are concerned, N can be determined, given the anode current. This is because I_a is unaffected by pulse overlap or any of the other sources that cause loss of counts in photon counting. The realization that I_a varies linearly with N is the basis of the method about to be described.

The instrumentation for determining dead time need not be sophisticated. The requirements are a random source of photons of adjustable rate, a PMT with a switchable load resistor (either $50\ \Omega$ or $100\ k\Omega$), and an amplifier discriminator of fixed threshold—typically with a pulse-pair resolution of about 20 ns, as measured with a $50\ \Omega$ pulse generator. The voltage divider should ideally be an active one to maintain linearity up to anode currents of about $100\ \mu\text{A}$. Alternatively, a divider with a high standing current could be used. Operating at negative HV facilitates the measurement of the anode current or, more conveniently, the voltage developed across a $100\ k\Omega$ load resistor may be measured and the corresponding count rate, n , for each anode current setting noted. The use of a battery-operated digital multimeter (DMM) floated across the load resistor provides a means for dealing with positive HV, but care must be exercised to avoid the obvious electrical hazard. Anode current is proportional to N for positive and negative HV since dead time does not apply.

The calibration method, devised at low count rates, relies on establishing a relationship between I_a and N . For example, a dead time of 20 ns and $n = 100\ \text{kHz}$, $n\tau = 2 \times 10^{-3}$ leads to a negligible correction in either type of $\sim 0.2\%$ (Note at a rate of $10\ \text{kHz}$ the correction is $\sim 0.02\%$ and truly negligible). Under the stated low count rate conditions, N and n are essentially equal and correspond to a known anode current that can be measured. At high count rates, N is deduced from the corresponding anode current. A typical plot is shown in Fig. 4.25, where the onset of non-linearity is noticeable beyond about 1 MHz. Count rate linearity is considerably improved by applying a Type II correction and assuming $\tau = 17\ \text{ns}$. The appropriate τ value is deduced by trial and error, with the proviso that the assumed value of τ does not lead to excessive super-linearity. Note that the dead time deduced with a pulse generator is 20 ns but a value smaller than this gives the optimum correction. This is because at low count rates the correction is relatively insensitive to the adopted τ value. For N extending to 20 MHz, there is an error of $< 5\%$ in the correction of n but at higher count rates there is a significant loss of counts. The two term correction of (4.107), assuming $\tau = 17\ \text{ns}$, is also shown in Fig. 4.25 where improved correction up to 30 MHz is evident. Figure 4.26 is an alternative and preferred presentation of the data in the previous figure. It is recommended, in the course of these measurements, to view the output of the

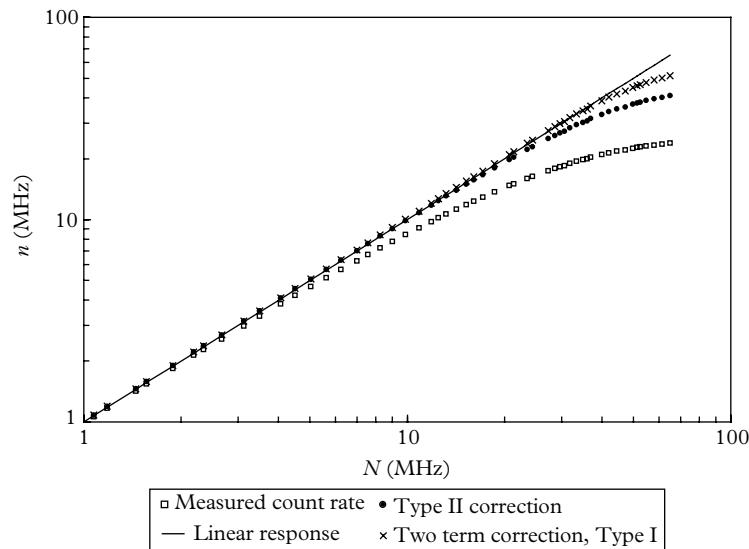


Fig. 4.25. Measured counts, n , as a function of true counts, N . The ‘best fit’ dead time correction of 17 ns lifts the measured counts closer to the straight line, but without overshooting.

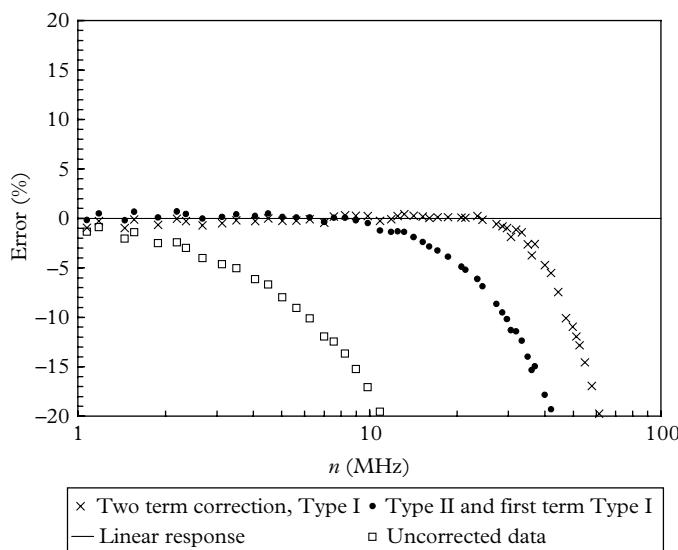


Fig. 4.26. Data is taken from Fig. 4.25, assuming $\tau = 17$ ns. Error is plotted as $(n - N)/N\%$ in order to enhance detail.

discriminator on an oscilloscope—both free running and single shot. Irregularities in performance, already alluded to, soon become apparent and perhaps alarming.

It should not be surprising, given the complex nature of dead time, to find that neither of the two formulae for Type I and Type II dead time transform all experimental points of a given set, (N, n) , on to a straight line. Linearization is easily accomplished where the dead time correction is small, but at high count rates both formulations under correct. The results in Fig. 4.26 are representative of performance with a typical threshold discriminator (either a nuclear-instrument-module (NIM) type or a self-constructed amplifier discriminator). Counting losses, of the order of 1 % in the raw data, begin to show at 1 MHz but, with Type II correction, this can be extended to ~ 10 MHz; the two term correction is effective up to 30 MHz. There is scope for enhancing the dynamic range by increasing the adopted value for dead time to, say, 20 ns; this has the effect of boosting counts in the high-count-rate domain, thus incurring a small degree of supre-linearity in the mid range. An alternative and arguably better way to correct for dead time is to construct a look-up table based on a calibration of the type shown in Fig. 4.26. In this way it is possible to measure count rates with an accuracy of $\sim 2\%$ at low rates, increasing to about 5% at 50MHz for a medium-fast PMT. In Fig. 4.26 it is clear that the two term Type I provides the best correction for the discriminator used in this study. It correctly predicts the turnover and hence ambiguity evident in Fig. 4.24. There are both an easy solution and a not-so-easy solution to this problem: the easy and sensible solution lies in measuring the anode current, which does not fold-over at high count rates, provided the PMT biasing arrangement holds. The difficult approach lies in detecting when the current flow from the high-voltage supply exceeds an appropriate level, indicative of count rate folding. It is feasible with active dividers to achieve linear operation, with respect to anode current and light level, for currents up to 300 μ A. PMT behaviour is uncertain beyond this point, depending upon the design details of the voltage divider and the individual PMT characteristics. Generally, a PMT suffers collapse in its output and the anode current ultimately decreases. This state of operation, brought on by high light levels, is potentially damaging to the PMT and any associated electronics—this mode of operation should be avoided. We have seen that dead time has a contribution, stemming from the finite bandwidth of the PMT. The half width of an output pulse initiated by a single photoelectron ranges from ~ 100 ps for a microchannel device to ~ 15 ns for a box-and-grid PMT. The statistics of random, but steady, light sources are such that a prevalence of short time intervals will always occur, thereby exacerbating pulse overlap. This form of dead time has a statistical nature and it is essentially indeterminate, since every anode pulse has a unique time signature—arguably, dead time is of the order of twice the fwhm of the anode signal. There is always some contribution to dead time from instrumentation used in conjunction with a PMT, which may be any one of the following: a discriminator, a single-channel analyser (SCA), an MCA, or a digitizer. These instruments have different dead times, with fast digitizers undoubtedly the least affected.

4.19.7 Suppressing correlated signals

Afterpulses and other correlated signals generally occur within 10 µs of an initiating event and they are more prevalent in the background than in signal. There are therefore advantages to be gained, particularly in low-light-level applications, by deliberately increasing the dead time of a counter. The consequential loss in counts is minimal in such arrangements but the unwanted contribution from correlated pulses may be significantly reduced. Note there is only a 1% count rate correction for $\tau = 10 \mu\text{s}$ at $n = 10^3$ counts/s, for example. This seemingly counter-intuitive mode of operation is most appropriate when signal and background counts are low, as in synchronous detection. The benefits derived from imposing additional dead time can be gauged by performing a series of multichannel scaling measurements on the same PMT, but under different dead time conditions. The result of taking 1000 background measurements, each of 1 s duration, with τ ranging from 30 ns to 1 ms, is shown in Table 4.4. The mean dark count with $\tau = 30 \text{ ns}$ is 8.1 counts/s as the PMT chosen is one of low background. The benefits derived from increasing dead time are reflected in the achievement of lower background counts. In addition, there is a reduction in the parameter ‘var/mean’, which is seen to approach its theoretical value of unity. The entries in Table 4.4 verify the assertion that most of the contributions to the variance are generated within 10 µs of the main pulse. Extending τ beyond this dead time setting brings further improvement, but increases the magnitude of the dead time correction. The effect of dead time on signal counts is similar to that observed with background, although not quite as striking.

A histogram of the distribution in counts confirms the origin of high variance contributions from the tail. This is shown in Fig. 4.27 for a dead time setting of 30 ns, together with a Poisson distribution of mean 5.54 (the last entry in Table 4.4). The impressive reduction in noise by deliberately adding dead time is highly beneficial in low-light-level applications. Measurements show that afterpulse rates are proportional to signal strength, but note with DC detection there is no means for reducing their contribution to noise.

The same data presented in Fig. 4.28 highlight the noise inherent in the background of this particular PMT. For a mean of 8 counts/s, the Poisson distribution predicts the occurrence of more than 20 counts in any of 1000 samples as

$$p(\geq 20) = \sum_{n=20}^{\infty} e^{-8} \times 8^n / n! \times 1000 = 0.25.$$

Therefore, a set of 1000 repeated readings, if properly behaved, is unlikely to contain a single reading greater than 20 counts, whereas Fig. 4.28 shows over 30 such occurrences in the set of measurements. Samples with excessive high counts disappear by imposing the dead times indicated in Fig. 4.28 (middle) and (bottom). Complete understanding of the origins of non-statistical counts in the background is patchy. Afterpulses generate two photoelectron events, and

Table 4.4 A reduction in dispersion, as a result of imposing additional dead time, may be gauged by the significant reduction in var/mean.

$\tau (\mu\text{s})$	Mean	var/mean
0.03	8.14	4.39
1.00	6.71	2.42
10.0	5.71	1.53
100	5.23	1.24
1000	5.54	1.05

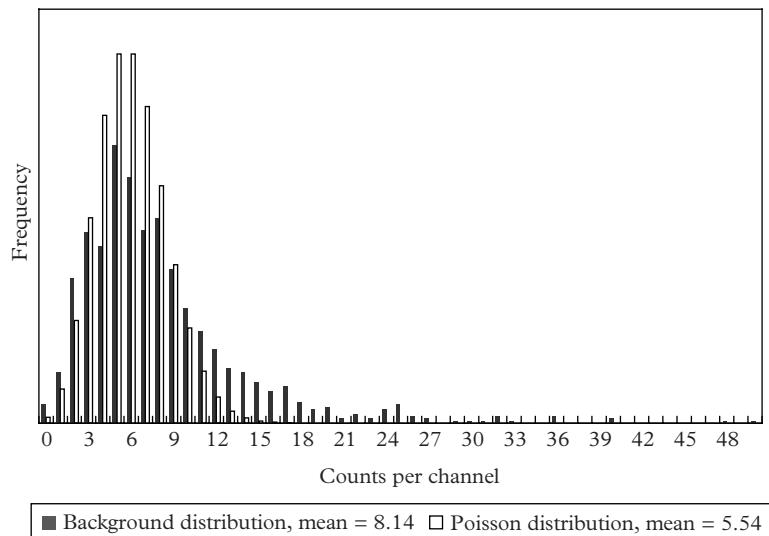
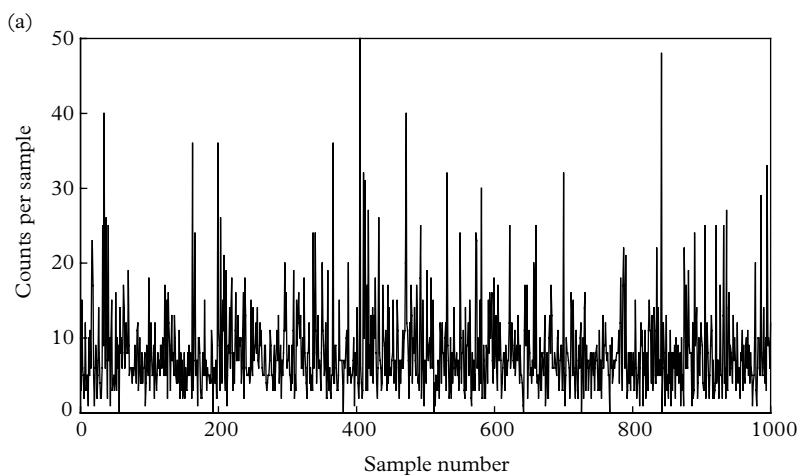


Fig. 4.27. Frequency distribution of dark count data for a PMT with a mean dark count rate of 8.14 counts/s, compared with a Poisson distribution of the same area. Note the excess in channel counts beyond Channel 10.



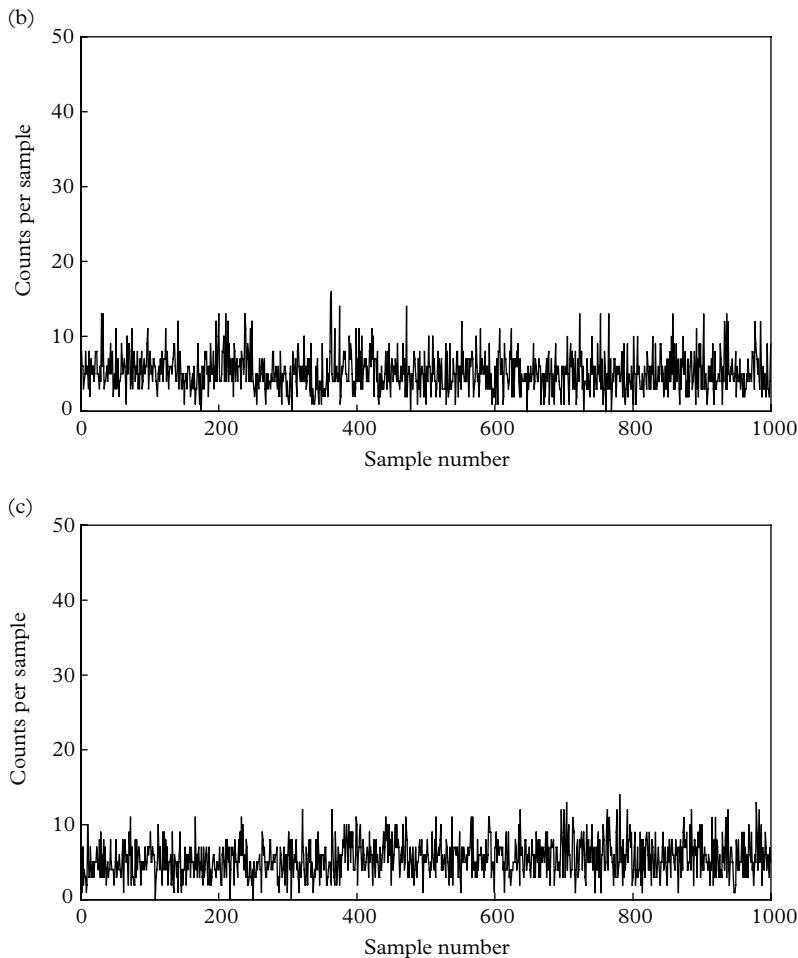


Fig. 4.28. (a) Repeated background measurements taken with a 9125B PMT and a dwell time of 1 s. The most probable count rate is 5.54 s^{-1} , and the mean rate is 8.14 s^{-1} . The dead time for these measurements was set to 30 ns.

(b) Background measured with an imposed $10 \mu\text{s}$ dead time. High count rate events have all but disappeared and there is an absence of counts exceeding 17 per sample.

(c) Background measured with the imposition of 1 ms dead time. In this case, there is complete absence of counts greater than 14 per sample, although there is only a small improvement over the scenario in the middle panel.

REFERENCES

- fluorescence in the window, caused by the passage of relativistic charged particles, may be the dominant source of enhanced count rates. The frequency of readings containing in excess of 20 pulses is compatible with the cosmic ray muon rate. This is about 1/10 s for this PMT type (based on the area of the window) and may explain the burst like nature of the background.
-
- Barletta et al. (eds). (2000). Second Conference on “New Developments in Photodetection” BEAUNE99: Beaune, France, June 21–25, 1999. *Nucl. Instr. and Meth. in Phys. Res. A*, **442**, 1–485.
- Barletta et al. (eds). (2003). Proceedings of the 3rd International Conference on New Developments in Photodetection, Beaune, France 17–21 June 2002. *Nucl. Instr. and Meth. in Phys. Res. A*, **504**, 1–390.
- Barletta et al. (eds). (2006). Proceedings of the 4th International Conference on New Developments in Photodetection, BEAUNE 2005, Fourth International Conference on New Developments in Photodetection, Beaune, France, 19–24 June 2005. *Nucl. Instr. and Meth. in Phys. Res. A*, **567**, 1–404.
- Evans, R. D. (1955). *The Atomic Nucleus* (2nd ed.). McGraw-Hill, New York.
- Feller, W. (1948). ‘On probability problems in the theory of counters. In *Courant Anniversary Volume*, Interscience, New York, 105–15.
- Jánossy, L. (1965). *Theory and Practice of the Evaluation of Measurements*. Clarendon Press, Oxford.
- Jones, R., Oliver, C. J. and Pike, E. R. (1971). Experimental and theoretical comparison of photon-counting and current measurements of light intensity. *Appl. Optic.*, **10**, 1673–89.
- Kelbert, M., Sazonov, I. and Wright, A. G. (2006). Exact expression for the variance of the photon emission process in scintillation counters. *Nucl. Instr. and Meth.*, **564**, 185–9.
- Knoll, G. F. (1989). *Radiation Detection and Measurement* (2nd ed.). John Wiley & Sons, Inc., Hoboken, NJ.
- Kosten, L. (1943). On the frequency distribution of the number discharges counted by a Geiger-Müller counter in a constant interval. *Physica*, **10**, 749–56.
- Lombard, F. J. and Martin, F. (1961). Statistics of electron multipliers. *Rev. Sci. Instr.*, **32**, 200–1.
- Morton, G. A., Smith, H. M. and Krall, H. R. (1968). Pulse-height resolution of high gain first dynode photomultipliers. *Appl. Phys. Lett.*, **13**, 356–7.
- Oliver, C. J. (1977). *A Comparison of Current Measurement with Photon Counting in the Use of Photomultiplier Tubes*. ET Enterprises Reprint Series, RP066. ET Enterprises Ltd, Uxbridge. Available at http://www.et-enterprises.com/files/file/technical-information/rp066_comparison%20current%20meas.%20photon%20counting%20in%20use%20of%20pmt.pdf (accessed 27 Jan 2017).

- Müller, J. W. (1973). Dead-time problems. *Nucl. Instr. and Meth.*, **112**, 47–57.
- Post, R. F. and Shiff, L. I. (1950). Statistical limitations on the resolving time of a scintillation counter. *Phys. Rev.*, **80**, 1113.
- Riley, K. F., Hobson, M. P. and Bence, S. J. (2002). *Mathematical Methods for Physics and Engineering* (2nd ed.). Cambridge University Press, Cambridge.
- Stein, J. *et al.* (1996). X-ray detectors with digitized preamplifiers. *Nucl. Instr. and Meth. in Phys. Res. A*, **113**, 141–5.
- Wright, A. G. (2007). The statistics of multi-photoelectron pulse-height distributions. *Nucl. Instr. and Meth. in Phys. Res. A*, **579**, 967–72.

5

Secondary emission
and gain

5.1 Introduction

High and relatively noiseless gain is one of the principal attributes of PMTs and it is this property, together with a bandwidth approaching 1 GHz, that distinguishes the vacuum PMT from other light detectors. It is important to be clear about what is meant by the term gain, g , where PMTs are concerned. It refers to the ratio of the output charge, q_a , at the anode, to an input charge of 1 pe, e , emanating from the photocathode. Quantitatively $g = q_a/e$ is the dimensionless multiplier gain but, to emphasize its statistical nature, it is generally denoted by the mean gain, $\langle g \rangle$. It is essential to distinguish $\langle g \rangle$ from a related dimensionless parameter, G , which refers to the response of a PMT to a steady cathode current, I_k ; G is defined as the ratio of the anode current to the photocurrent, I_a/I_k . It might appear that the two definitions should lead to the same numerical figure: with $\langle g \rangle$ the ratio of charges, and G the ratio of currents. However, G is always less than $\langle g \rangle$ because of two considerations: a fraction of photoelectrons fail collection at the first dynode (landing on the focussing structure, for example); second, a fraction $p(0)$ of photoelectrons incident on d_1 fail to propagate as is to be expected on statistical grounds (discussed in 4.18.4). These loss mechanisms do not affect $\langle g \rangle$ since it is based only on those photoelectrons that propagate. This aspect of PMT performance also concerns the controversial subject of collection efficiency, discussed in Chapter 10.

Manufacturers specify sensitivity in various ways. The unit of amps per lumen signifies how many amps would be measured at the output of a PMT for a light input of one lumen, and manufacturers provide the overall HV required to attain a quoted PMT gain, G . The origin of this parameter refers to a time during which PMT applications were predominantly concerned with white light sources. The advent of lasers saw the adoption of the more appropriate unit of amps per watt and, where scintillation crystals are used, picocoulombs per MeV is preferred. In these three cases, the concept of gain has been extended to include other elements of the detection system. Parameters with dimensions are actually indicators of sensitivity, but not gain, although few users appreciate this subtlety. As a generalization, $\langle g \rangle$ is usually the preferred parameter for academic and scientific purposes, whereas G is of practical relevance in industrial applications of PMTs. PMT sensitivity is covered in detail in 2.3 but, in this chapter, most references to gain refer to multiplier gain, unless otherwise stated.

Gain is achieved in vacuum PMTs through the consecutive action of a multiplicity of electrodes, called dynodes. Their number ranges from 6 to 14, and the appropriate selection of the number depends on the overall gain required. The set of dynodes within a given PMT includes some of a different shape. Common dynodes, as the term suggests, share the same shape and gain characteristic. For example, the first, second, and last dynodes in a linear focussed multiplier are different from the remainder (see 5.6). The gain of the first dynode is usually considerably higher

than that of a common dynode. There are special PMTs with only one, two, or three dynodes, developed for particular high-energy physics experiments where light levels are high. On the other hand, devices without a photocathode are intended for ion or electron detection and may have up to 21 dynodes. Some PMTs use a continuous form of electron multiplier, such as a single Channeltron™ or microchannel plate. Both these devices derive gain from secondary emission. There are other detectors in which gain is realized through electron-hole generation: in a silicon diode (HPDs); in an avalanche diode (SiPMs); or through gas multiplication (gas electron multipliers (GEMs)). The focus of this chapter will be on the phenomenon of secondary emission and the underlying physics of the process. The statistical aspects of dynode behaviour are covered in greater detail in Chapter 4.

5.2 Gain generation with discrete dynodes

This section concerns the relationship between overall gain, $\langle g \rangle$, the number of stages, k , and their individual gains, δ , where δ is the ratio of the charge leaving a dynode to the charge incident on that dynode. The electron gain that can be realized by a system of k discrete dynodes is based on a multiplication process. Assuming that all k stages have the same stage gain, δ , the overall gain, $\langle g \rangle$, is

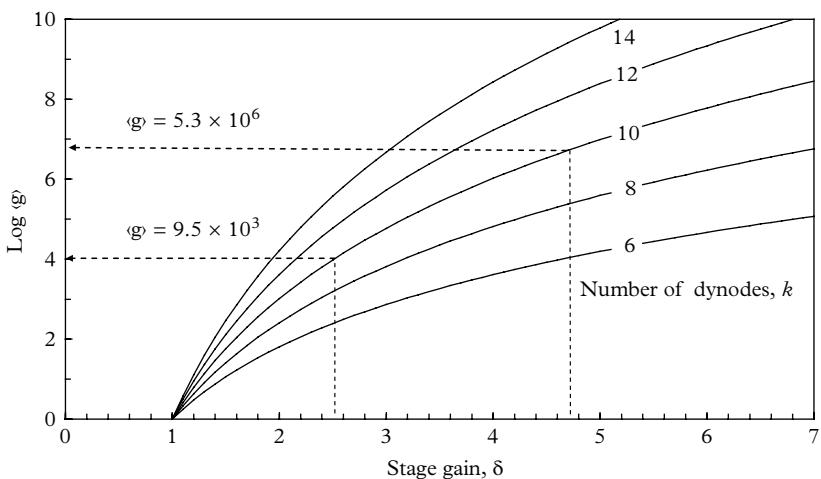


Fig. 5.1. The relationship between δ , k , and $\langle g \rangle$ derived from (5.1) to (5.3). The gain in most PMT applications lies between 10^4 and 10^8 . The dashed lines refer to the overall gain at the inter-dynode voltage of 100 V, taken from Fig. 5.2. It should now be obvious why manufacturers invariably present gain data using a log scale for the ordinate.

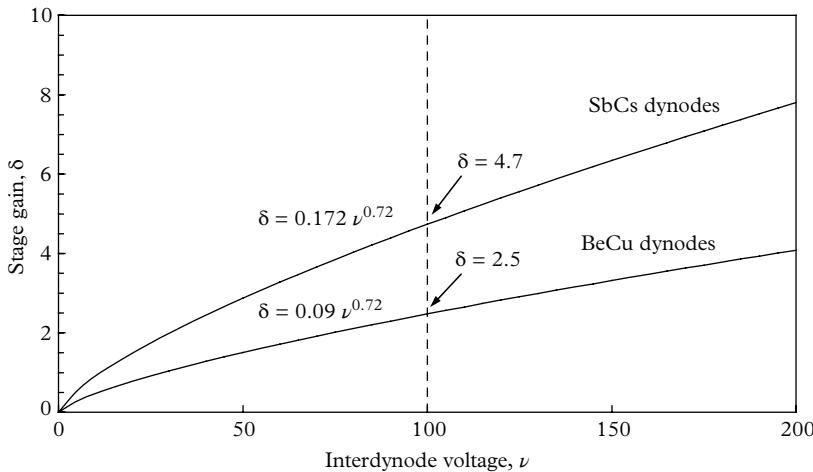


Fig. 5.2. Typical stage gain for SbCs and BeCu secondary emitting surfaces, based on $\delta = a\nu^\alpha$. Measurements made on a range of PMTs with BeCu and SbCs dynodes show that the index, α , is similar for the two materials, and the differences in stage gain derive from the numerical values for the coefficient a .

simply δ^k . Plotting this relationship reveals the range of combinations of δ and k that may be chosen to attain any specified overall gain. The parametric curves in Fig. 5.1 also show (1) the sensitivity of $\langle g \rangle$ to small changes in δ , and (2) the rapid increase in $\langle g \rangle$ with k . An appreciation of both these properties is essential in choosing, and in using, PMTs. Up to this point, no assumptions regarding the voltage dependence of δ have been made but, once this is given, as in Fig. 5.2, for example, we have a method for attaining the optimal choice of k for any required gain. In Fig. 5.2, for example, a PMT with ten BeCu stages, operating at 100 V/stage, will have an overall gain of $(2.5)^{10} = 9.5 \times 10^3$, at $V = 1100$ V, noting that a k -stage multiplier needs a voltage divider with $(k+1)$ stages. The gain under the same conditions for a multiplier with SbCs dynodes is $(4.7)^{10} = 5.3 \times 10^6$, and the way in which this is determined is shown in Fig. 5.2.

A semi-empirical formula relating the secondary emission yield, as a function of primary electron energy, can be developed as follows: in the absence of inelastic scattering losses, a linear relationship between the inter-dynode voltage, ν , and the mean number of secondaries, δ , may be anticipated, in the form of $\delta = a\nu$. Allowance for scattering losses follows by assuming an index for ν , less than unity, leading to

$$\delta = a\nu^\alpha. \quad (5.1)$$

Equation (5.1) is superimposed on experimental points in Fig. 5.3, for the first dynode. For a multiplier with k stages, all of different gain,

$$\delta^k = a_1 v^{\alpha_1} \times a_2 v^{\alpha_2} \times \cdots \times a_k v^{\alpha_k}, \\ \langle g \rangle = (a_1 \times a_2 \times \cdots \times a_k) \times v^{\alpha_1 + \alpha_2 + \cdots + \alpha_k}.$$

For a uniform inter-dynode voltage distribution, and voltage V ,

$$\nu = \frac{V}{k+1}, \\ \delta^k = (a_1 \times a_2 \times \cdots \times a_k) \times \left(\frac{V}{k+1} \right)^{\alpha_1 + \alpha_2 + \cdots + \alpha_k}. \quad (5.2)$$

and, if the dynodes are identical in their gain characteristics,

5

$$\delta^k = \frac{a^k V^\beta}{(k+1)^\beta}, \\ \langle g \rangle = \delta^k = c V^\beta, \quad (5.3)$$

where $\beta = k\alpha$, and c is a constant. If stage gains differ in exponent and coefficient, equation (5.3) still applies, with $\beta = \alpha_1 + \alpha_2 + \cdots + \alpha_k$, and $c = (a_1 \times a_2 \times \cdots \times a_k) / (k+1)^\beta$. A gain–voltage relationship can always be constructed, given at least two coordinate pairs (V, g) . These relationships are of fundamental importance in PMT applications, in particular voltage divider design, and their validity is confirmed in Figs 5.3 and 5.4.

From Fig. 5.4, the relationship between $\langle g \rangle$ and V is

$$\log \langle g \rangle = 7.88 \log V - 16.88, \quad (5.4)$$

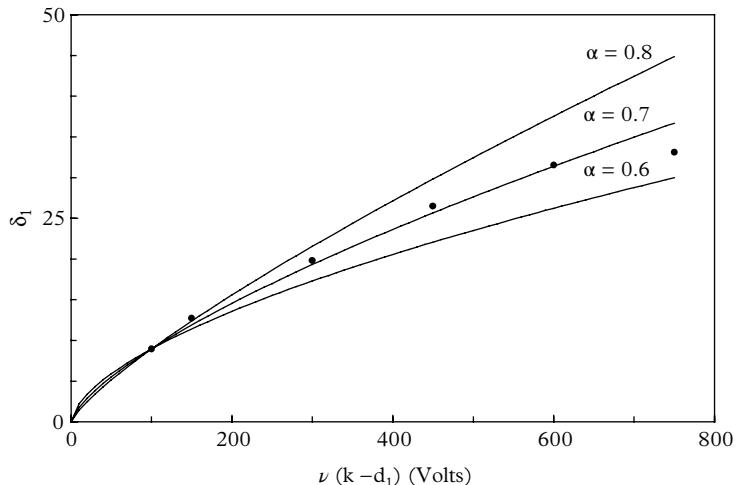
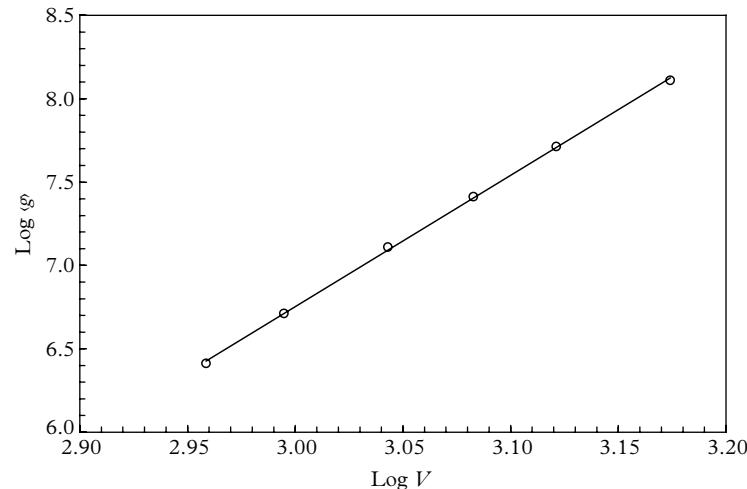


Fig. 5.3. Measured gain of the first dynode (\bullet) of a high-gain linear focussed PMT, compared with the predictions of (5.1) for $\alpha = 0.36$, and the α values indicated.



5

Fig. 5.4. An example of a gain–voltage characteristic for an 11-stage PMT. The voltage divider is uniform with all resistors of equal value. Such dividers are sometimes called ‘linear’, which may cause confusion where gain linearity is under discussion. Circles represent measured data, while the straight line is the relationship in (5.4) and also expressed as a power relationship in (5.5).

which is more conveniently expressed as a power relationship:

$$\langle g \rangle = cV^\beta, \quad (5.5)$$

with $c = 10^{-16.88}$, and $\beta = 7.88$, for V measured in volts.

On the assumption that all stages are identical, and given a gain–voltage relationship as in Fig. 5.4, we can work in reverse to establish δ in terms of v :

$$\delta = 0.172v^{0.72}. \quad (5.6)$$

Note (5.5) and (5.6) should be regarded as representative functional relationships, since individual stage gains in a given PMT will vary from one to another. The required gain can only be specified with knowledge of the application but, for the moment, consideration is given to one of the more demanding applications—that of photon counting. Based on the previous example, is the gain of 5.3×10^6 for an 11-stage PMT sufficient to detect events initiated by single photons? At this gain, the output pulses contain an average charge of $1.6 \times 10^{-19} \times 5.3 \times 10^6 \sim 1.0$ pC and, if this charge were delivered in, say, 10 ns to a 50Ω load, the peak voltage would be of the order of -4 mV, and just discernible on an oscilloscope. With this straightforward analysis it is thus possible to explain why practical photon counting is performed at a gain in the range $5 \times 10^6 < g < 2 \times 10^7$. This rough calculation, applicable to both DC and pulsed currents, may be applied to all applications in which the minimum input charge (or current) is known. The

considerations so far relate to mean gain in order to highlight the fact that PMTs follow a power law of the form $g_p = cV^\beta$. An important point to note is the rapid increase in gain with applied HV, because of the power law.

5

5.3 Single-photoelectron noise spectra

Illuminating the photocathode with single photons produces amplified signals at the anode, with a spread in charge about some mean value. A spectrum, in the form of a pulse height distribution, accumulated by an MCA, for example, shows a spread in pulse heights and hence in gain from one detection to the next. The term SER refers to the pulse height distribution at the anode, ideally containing sufficient events to provide a smooth curve. The multiplier assigns a different gain to every detected photon. It is said to be noisy, and the performance of a PMT depends critically on the quality of its SER. The technique for recording such a spectrum is explained in 4.1.1. A representative sample of SERs for vacuum PMTs is shown in Figs 5.5 to 5.7 to illustrate the widespread nature of this parameter. It can be seen immediately that the distributions are

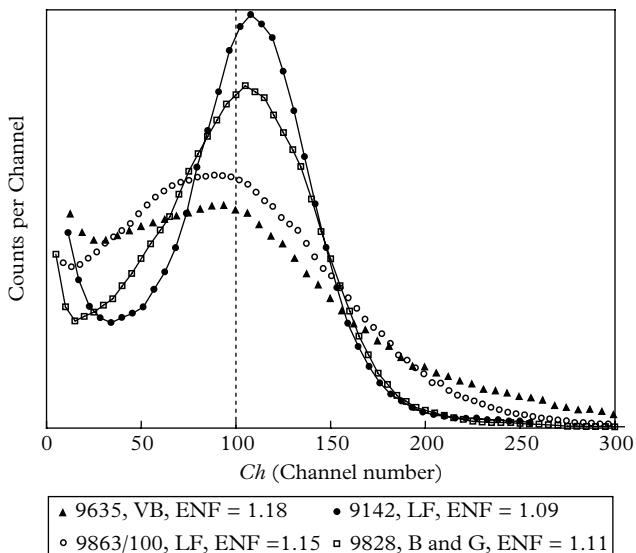


Fig. 5.5. Pulse height distributions for single-photon excitation (450 nm wavelength) distributed evenly over the photocathode area. Each distribution refers to the particular dynode type indicated in the legend. All curves have been normalized in area, and to the same mean pulse height of Channel 100. Note that the mean, $\langle Ch \rangle$, and the peak channel are not coincident. Channels can be related to gain by calibration.

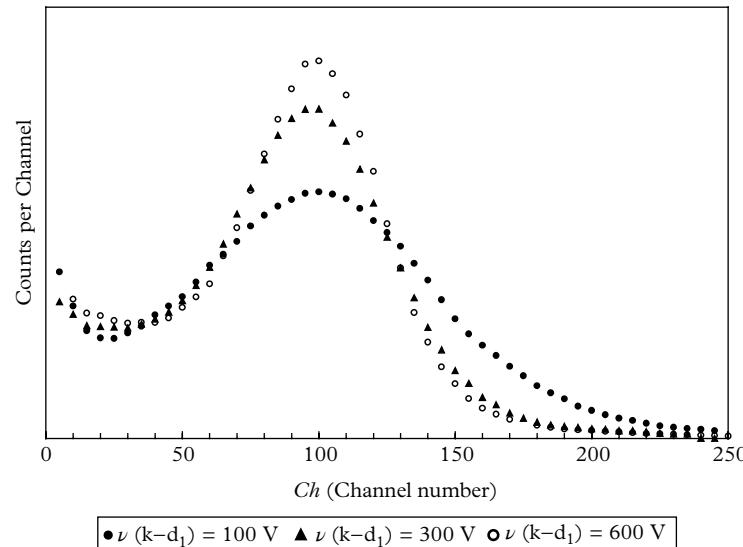


Fig. 5.6. The effect of changing $\nu(k-d_1)$ on the SER of a linear focussed PMT of high d_1 gain. Illumination conditions are the same as in Fig. 5.5. The area under each curve is normalized, and the peaks are set to Channel 100.

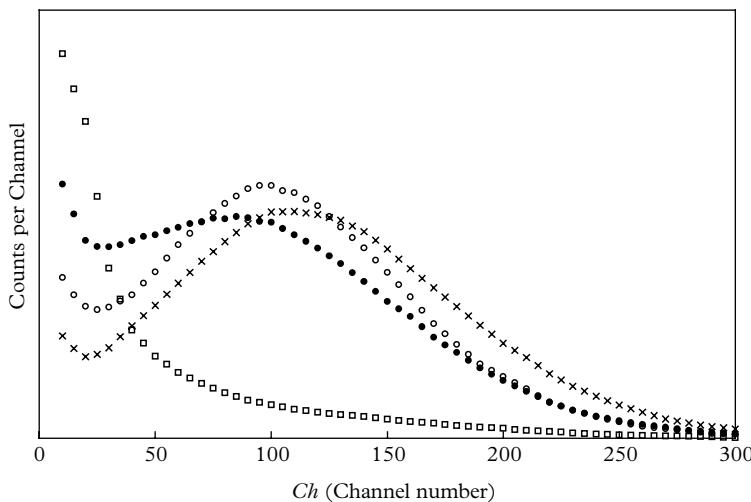


Fig. 5.7. Showing that the SER of a Venetian blind PMT is sensitive to the location of photodetection and to the photoelectron arrival directions; ●, uniform illumination of the photocathode; ○, central 2 mm diameter spot on the axis of the PMT; ✕, photoelectrons incident normal to the plane of the slats; □, photoelectrons incident parallel to the orientation of the d_1 slats corresponding to trajectory (a) in Figs 5.20 and 5.21.

non-symmetrical about their peak values, because of an obvious excess of counts at low pulse heights. Variance is calculated for the distributions, using standard numerical procedures, and then substituted into the appropriate formulation for noise (see Chapter 4, (4.77) and (4.78)). An excess noise factor, ENF, introduced in 4.10.3, gives a measure of the resolution quality for pulsed applications, and, likewise for DC applications, imposes noise on a measured signal level. Noise factor computations, based on the variance figure for the distributions, range from 1.09 to 1.18. The abscissa is labelled ‘Channel number’, although it is readily expressed in terms of gain through calibration. Note that distributions are asymmetric about the peak position because of an excess of events extending from Channel 1 to Channel 50. Pulses within this area contribute significantly to the variance, and the peak-to-valley ratio is thus an alternative figure of merit for an SER.

Figure 5.6 shows the effect of increasing the $v(k-d_1)$ voltage and hence the gain of the first dynode. Improved resolution follows but the small signal region is hardly affected. Figure 5.7 reveals the positional dependence, relative to the photocathode, of the SER using a small focussed spot. This highlights a particular deficiency of Venetian blind multipliers, which is examined in detail in 5.6.1. In the sections that follow, we explore theoretical considerations to explain the behaviour of multipliers, starting with the secondary emission process.

5.4 Secondary-electron emission

5.4.1 Classical description

Electrons acquire kinetic energy following acceleration in an electric field and, if subsequently brought to rest by impact on a material surface, they will lose this energy through ionization and excitation. These two processes are responsible for producing electrons and lead to the generation of gain. There are three types of material to consider when discussing secondary emission: metals, semiconductors, and insulators. They have different secondary emission properties from one another but all three are utilized in PMTs. Secondary emission from any part of the internal structure, other than the dynodes, is clearly undesirable, and such phenomena are discussed in 5.5. According to classical theory, the mechanism of emission has three stages: the production of electrons within the solid; their migration; and the escape from the solid–vacuum interface. An electron multiplier comprises a set of dynodes, the active surfaces of which consist of a thin layer of secondary emissive material. The dynode potentials are arranged in a monotonic positive sequence, applied between the first and last member of the set. These inter-dynode potentials provide the electric field gradients to release and accelerate electrons from one dynode to the next, and the kinetic energy acquired in this process is released on impact for the creation of secondary electrons. Those electrons

generated by a particular dynode become the primary electrons for the next, resulting in a cascade for which the attainable gain depends upon the number of stages, the magnitude of the applied voltage, the nature of the secondary emission surface, and the geometry of the particular dynodes.

Although secondary emission is a quantum mechanical effect, the classical treatment can explain these aspects of the process relevant to PMTs—and it is also easily understood. The study of secondary emission has a long history originated by Franck and Hertz, some 100 years ago. There is considerable published work but the comprehensive review by Dekker (1958), despite its antiquity, remains a worthy paper for study. Another review paper from an even earlier date is that of Allen (1950), one of the pioneers in the early use of PMTs: ‘Recent applications of electron multiplier tubes’, as the title suggests, places emphasis on the nature of secondary emission from materials relevant to PMTs. There is a small number of additional papers that are specific to PMTs: notably, one by Simon and Williams (1968) deals with GaP, a new dynode material at that time. Coates (1970a, 1971) produced a set of papers covering detailed investigations on PMT background, collection efficiency, and secondary emission. The focus of his investigations was the RCA 8850, a PMT that includes a GaP first dynode of high gain.

The primary concern centres on two relationships: the energy distribution of the secondary electrons, and the number of secondaries, δ , produced per incident primary. Figure 5.8 embodies the essential features of an experimentally determined secondary-electron energy spectrum for metals, semiconductors, and insulators. Examples of such spectra may be found in the technical literature, all of

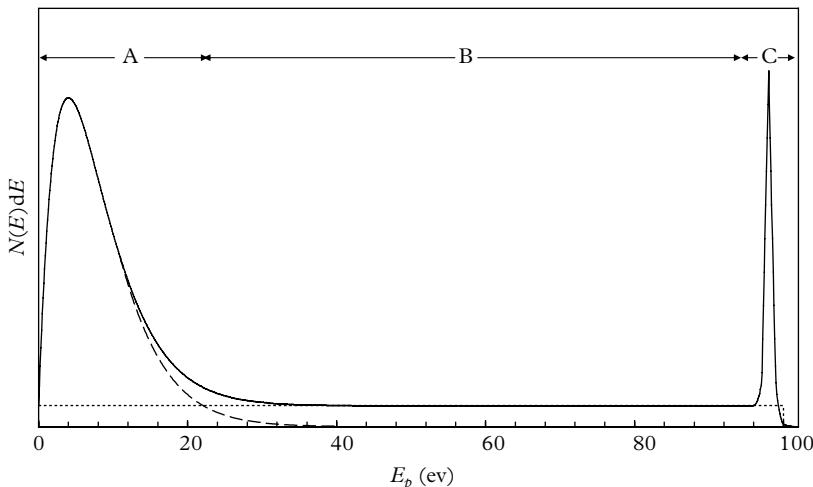


Fig. 5.8. The energy distribution of electrons emitted by a surface bombarded by primary electrons of energy $E_p = 100$ eV. The processes involved in the three regions A, B, and C are explained in the text, and in the diagrams of Fig. 5.9.

which conform to the general shape depicted in Fig. 5.8. A distinction is made between three categories of electrons: true secondaries, inelastically scattered electrons, and elastically scattered electrons. This categorization is based on the different ways in which free electrons may arise, and the features in the secondary-electron spectrum of Fig. 5.8 are explained in terms of this classification:

Region A: Comprises true secondaries that, by definition, are the released electrons which originally occupied bound states within the layer.

Region B: A primary electron may, after creating several true secondaries, suffer an inelastic scatter at any stage in the process (also called backscattering) carrying with it a portion E'_p of E_p , with E'_p in the range $0 < E'_p < E_p$. The distinction between true secondaries and backscattered electrons is necessarily vague, particularly in the low-energy region of the spectrum, where the two means of generation become indistinguishable. A noteworthy feature of B is its flatness over the entire energy range, with k , the proportion of primaries that suffer backscatter, ranging from 0.1 to about 0.3, depending primarily on the effective Z of the target material, and on E_p .

Region C: This includes all primaries that are elastically scattered by a process wherein they carry away the full primary energy, E_p . Also known as reflected primaries, they account for only a few per cent of events in the spectrum.

The diagram in Fig. 5.9 illustrates the ways in which contributions to A, B, and C are generated. The area defined by A accounts for 70 to 80% of the total number

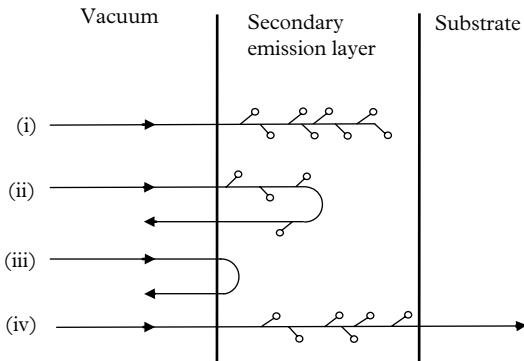


Fig. 5.9. Illustrative examples of the way in which a primary electron of energy E_p interacts with a secondary emissive layer. (i) The primary electron produces eight electrons, some of which escape the surface as true secondaries. (ii) The incident electron is inelastically scattered after creating four free electrons, some of which escape as true secondaries. (iii) The incident electron is elastically scattered from the surface of the layer. (iv) The incident electron has sufficient initial energy to traverse the secondary emissive layer into the substrate, where there is little chance of generating or collecting secondaries.

of secondaries produced, making the process indicated by (i) the dominant one. The entire energy of the primary electron is given to the production of low-energy electrons, with those produced sufficiently near the vacuum interface having a high probability of escaping the surface. The remainder are lost from the multiplication process. In inelastic scattering, indicated by (ii), the primary electron enters the surface, produces three electrons and is then backscattered, generating one further electron on exit. In addition to the high-energy, inelastically scattered electron, four low-energy secondaries may be emitted in this example. In elastic scattering, indicated by (iii), the primary is reflected without energy loss, and with a Lambertian distribution in angle. The secondary emission yield, δ , is a function of E_p , and there exists an energy, $E_p(\max)$, beyond which the number of secondaries decreases with increasing E_p . Three causes are identified for this case, which is indicated by (iv). First, the energy loss of the primary, dE_p/dx , varies approximately as E_p^{-1} (the Waddington law). Second, the probability of escape diminishes for those secondaries produced deep below the surface. Third, sufficiently energetic primaries may even propagate beyond the secondary emissive surface and into the supporting substrate, usually of low δ .

The secondary yield for smooth surfaces increases with the incident angle of the primary. This is understood qualitatively in terms of the production of a higher proportion of secondaries disposed near to the surface. Consequently, the secondary emission yield, as a function of incidence angle, θ , varies approximately as $1/\cos\theta$, which has been verified with enhancements of up to a factor of 2 at grazing angles reported by Allen (1950) and others. Dekker (1958) gives curves for smooth nickel carbide to illustrate this effect, contrasting the results for soot, for which there is no angular dependence. The explanation for the behaviour of porous surfaces is that some electrons become trapped. Both of these considerations have relevance to PMTs: the first dynode tends to have a considerably higher gain than the others in the multiplier, and one of the contributory factors to this enhanced gain is the angle of incidence of the photoelectrons.

Metals are characterized by δ values close to unity while, for insulators, such as the alkali halides, ceramics, BeO, and MgO, δ ranges from 6 to 20, depending on preparation methods. Insulators, which are used throughout the construction of a PMT, may acquire a surface charge upon bombardment by electrons; this will occur if their secondary emission coefficient is greater than unity. There may be serious effects on the apparent gain stability of a PMT subjected to changing light conditions, with undesirable practical consequences. Typical operating values for δ are 10 for Cs_3Sb , 20 for oxidized $\text{K}_2\text{CsSb(O)}$, and 40 for GaP(Cs). It should be borne in mind that the secondary-electron yield depends on how the sample is activated (e.g. via oxidation or heat treatment), vacuum conditions, and on the passage of time. Allen (1950) shows the effect on unprocessed BeCu upon heating, followed by combined heating and oxygen treatment. The secondary emission coefficient, initially only 2, is increased to 5, and finally attains a value

of 6. The need to reactivate BeCu multipliers, following exposure to air, is well known in the field of mass spectrometry. It might appear from the foregoing that gain is inherently unstable, given the dependence on the repeatability of chemical processes together with the magnifying action of the gain–voltage power law relationship of Fig. 5.4. However, a combination of improved manufacturing processes, including a burn-in at elevated temperature while operational (now a standard procedure), ensures stable gain shortly after switch-on. The evidence for these statements may be found in the numerous high-energy and astrophysics experiments involving the continuous operation of large numbers of PMTs, for many years and under low-light-level conditions.

5

5.4.2 Secondary-electron yield

We adopt the methodology of Dekker (1958), as adapted by Simon and Williams (1968), in predicting the yield curve for GaP, but the type of material is incidental. It is assumed that the secondary emission yield, as a function of E_p , can be written in the following form:

$$\delta = \int n(x, E_p) f(x) dx,$$

where $n(x, E_p)$ is the number of electrons produced at a distance x from the surface, and $f(x)$ is the probability that an electron produced at x is emitted into the vacuum. If e is the energy required to produce a secondary electron, then we can relate $n(x, E_p)$ to $-dE/dx$, the energy loss per unit path length of the primary electron, through

$$n(x, E_p) = -\frac{1}{e} \frac{dE}{dx}.$$

Making the reasonable assumption that $f(x)$ follows an exponential law of the form

$$f(x) = B \times \exp(-x/L),$$

where B is a coefficient that takes account of two physical processes: first, only a fraction of the excited electrons diffuse towards the surface; second, only a portion of these electrons escape the surface. Here L is the mean free path for the absorption of the excited electrons during their diffusion to the surface—effectively the escape depth. Using the experimental results of J Young (1956), the range of primary electrons, R , in a material of density, ρ , is given by

$$R = \frac{1.15 \times 10^{-5}}{\rho} E_p^{1.35} \text{ cm},$$

where R is in centimetres, and ρ in grams per cubic centimetre, but note that Simon and Williams (1968) erroneously use 10^{-6} instead of 10^{-5} in this formula. If we furthermore assume that $-dE/dx$ is constant throughout the primary electron range R , so that $-dE/dx = E_p/R$, we have, finally,

$$\begin{aligned}\delta &= \int \frac{B}{\epsilon} \times \frac{E_p}{R} \times \exp(-x/L) dx, \\ &= \frac{BE_p}{\epsilon} L \left(1 - \exp(-R/L)\right).\end{aligned}\quad (5.7)$$

5.4.3 Theory and measurement

A normalized curve for δ/δ_{\max} versus $E_p/E_p(\max)$ is easily derived from (5.7); it is found that experimental results, for different materials, tend to follow this common so-called reduced yield curve. Such relationships are given by Dekker (1958) and by Simon and Williams (1968), among others. An example of a reduced yield curve, together with experimental data, is shown in Fig. 5.10 for MgO, Ge, and Pt, supporting the proposition of a universal curve for all materials.

Following Simon and Williams (1968), we can plot (5.7) for GaP, for example, by assuming the following values: $\rho = 5.35 \text{ g/cm}^3$; $\epsilon = 8.7 \times 10^{-3} \text{ keV}$; $B = 0.5$; and $L = 2 \times 10^{-5} \text{ cm}$ (200 nm). The result is shown as the uppermost curve in Fig. 5.11. Two further curves for $L = 10^{-6}$ (10 nm) and 3×10^{-6} (30 nm) are also shown, to reveal the sensitivity of the relationship to this parameter.

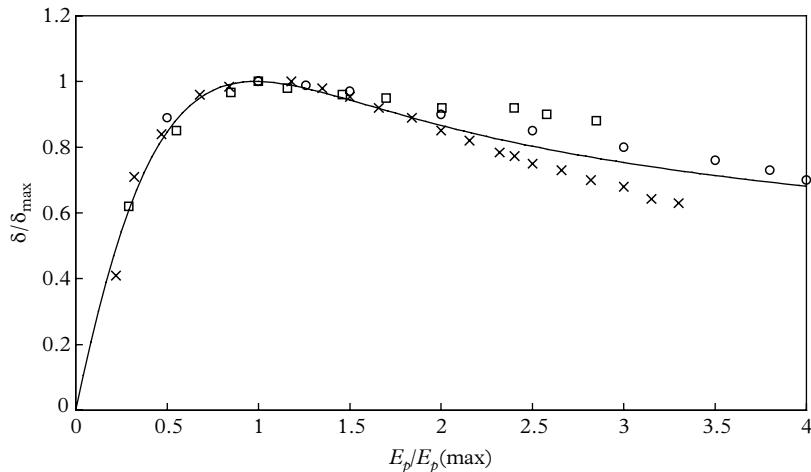


Fig. 5.10. Equation (5.7) plotted in normalized coordinates reveals that all secondary emission materials used in the reflected mode follow a universal curve. The same parameters adopted for Fig. 5.8 apply to the curve shown. The experimental data is from Simon and Williams (1968) for MgO (x), Ge (o), and Pt (□).

Simon, R. E. and Williams, B. F. (1968). Secondary-electron emission. *IEEE Trans NS15*, 166–70. Reprinted with permission from IEEE. ©1968.

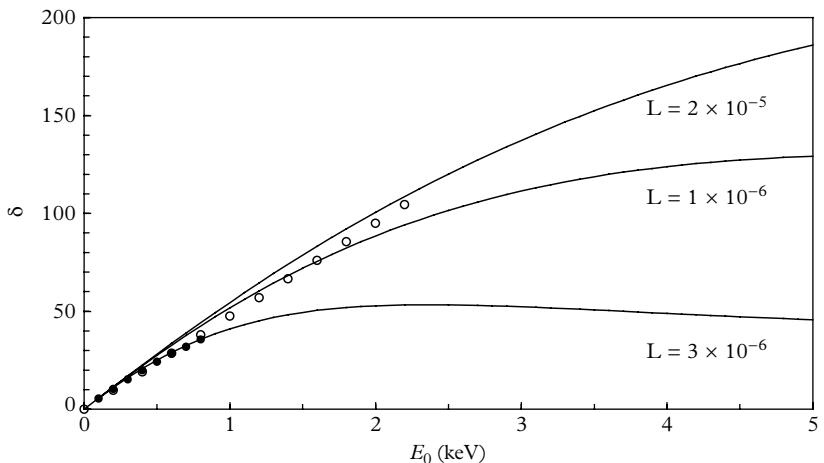


Fig. 5.11. Secondary emission yield, δ , as a function of primary energy for caesiated GaP computed from (5.7). Note the sensitivity to the assumed value for L , the mean free path. The best fit to experimental results is obtained with $L = 2 \times 10^{-5}$ cm for those of Simon and Williams (1968; ○), and 3×10^{-6} for the measurements of Coates (1970b; ●).

Measured secondary-electron yield curves for the more common dynode materials are given in Fig. 5.12, taken from Sommer (1972), Simon and Williams (1968), and the present author. Experimental points in Fig. 5.12 were determined by the method outlined in Fig. 5.16.

5.4.4 Secondary-electron statistics

The requirement that the number of secondaries always be an integer suggests that the distribution of electrons at the first dynode might follow Poisson statistics. Application of cascade theory in 4.10.4 predicts that the distribution at the anode is barely affected by dispersion introduced by the remaining dynodes. Hence, we need only consider the behaviour at the first dynode. This leads to the expectation that the output pulses from a multiplier, when excited by single photoelectrons, will result in a statistically well-defined distribution of φ electrons. The spread in pulse heights at the anode should approach a normal distribution, given the typical stage gains that apply to modern PMTs. However, compared with a Poisson distribution, what is generally observed is a peaked one but with an excess of small pulses. This composite distribution stems from two statistical sources, and the treatment in 5.4.2 needs to be extended by including inelastic scattered electrons in the signal-generating process.

Inelastic scattering in Region B of Fig. 5.8 is particularly undesirable in any electron multiplier. Such electrons carry away energy that would otherwise be

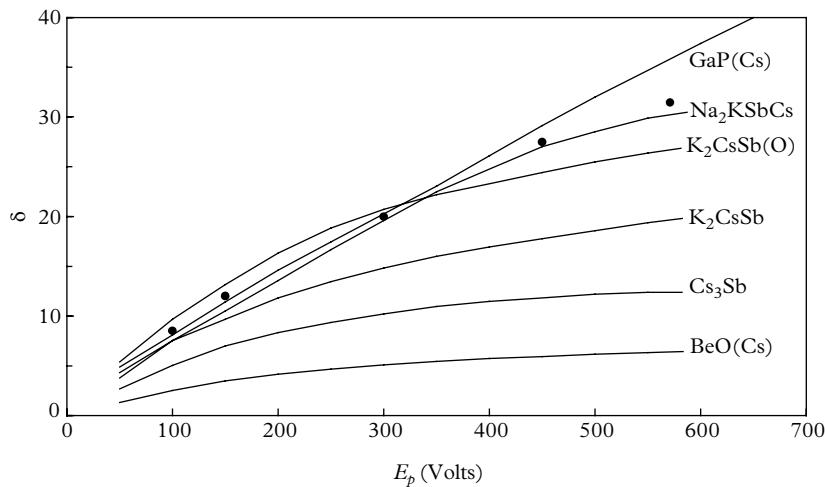


Fig. 5.12. The secondary-electron yield for various dynode surfaces. The experimental points (●) were taken by the author for a PMT in which the photocathode and the first dynode layers have the same chemical composition, namely, K₂SbCs. The lowest curve in the set refers to BeCu dynodes after oxidation and activation with caesium.

Sommer, A. (1972). Bialkali (K₂CsSb) photocathodes as a high gain secondary electron emitter. *J. Appl. Phys.*, 43, No. 5, 2479–2480. Reprinted with permission from AIP© (1972).

expended in producing true secondaries. Consequently, there is a reduction in yield, but the primary concern is the statistical one. There is a considerable literature on backscatter, although the majority of it concerns higher energies than those that apply to PMTs. The primary energies of interest in PMTs range from 10 to about 1000 eV, and the comprehensive and useful summary by Darlington (1975), for example, covers the range 10 to 100 keV. Nevertheless, the relationships between primary energy, scattered energy, angle of incidence, and angle of reflection, and their dependency on Z , can be inferred at lower energies. True secondaries and the backscattered electrons are emitted from the surface with a cosine distribution of escape angle (Darlington 1975). A small proportion of those backscattered electrons, with a favourable emission direction, land on the second dynode, making a contribution to the overall gain, and some return to d_1 following a circuitous path. The proportion of primary electrons that are backscattered, k , is a function of Z_{eff} , the effective atomic number of the material, defined by

$$Z_{\text{eff}} = \frac{mZ_A^2 + nZ_B^2 + \dots}{mZ_A + nZ_B + \dots},$$

for a compound $A_mB_n\dots$

The Z_{eff} values for the more common dynode surfaces are as follows: BeO, 6.7; GaP, 26; Na₂KSB, 34; K₂CsSb, 44; and Cs₃Sb, 54. The work of Thomas and

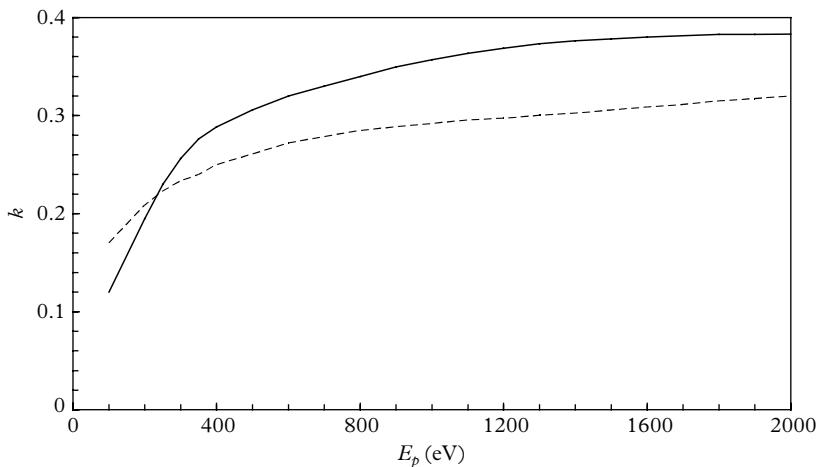


Fig. 5.13. Curves due to Stuchinsky (1970) for the backscatter coefficient k , for Cs_3Sb (solid curve) and $\text{Na}_2\text{KSB}(\text{Cs})$ (dashed curve).

Patterson (1970) for aluminium shows that k is independent of E_p for energies greater than 100 eV but decreases rapidly at lower energies. Stuchinsky (1970) has measured k for materials of direct relevance to PMTs, and he provides curves for k , starting at $E_p = 100$ volts. His data in Fig. 5.13 show the Z dependence very clearly, and we can assume that the curve for the bialkali K_2CsSb secondary emitter will lie intermediate to the curves shown for Cs_3Sb and $\text{Na}_2\text{KSB}(\text{Cs})$.

5.4.5 Temperature dependence of gain

While many of the physical properties of secondary emitters were exhaustively studied in the 1950s, data on temperature dependence are limited. A possible explanation for this lies in the magnitude of the temperature coefficient: it is so small that measurements need to be made over a significant temperature range to generate a measurable change in secondary emission. Dekker (1958) quotes a reduction of 10% in δ over a positive temperature range of 25°C to 500°C for germanium: that is, $-0.02\text{}/^\circ\text{C}$. For an insulator, MgO , the figure is $-0.017\text{}/^\circ\text{C}$ over a similar temperature span. Coefficients of the same order as these may be anticipated for activated dynodes within a PMT. Measurements made by the author on cooling PMTs indicate that δ increases by $\sim 0.2\text{}/^\circ\text{C}$ on cooling between -60°C and $+20^\circ\text{C}$. At temperatures below -60°C , there is an increase in the loss of gain to $0.4\text{}/^\circ\text{C}$ with cooling. Note these figures are for 12 stages of gain, so the temperature coefficient per stage is derived by scaling these numbers by a factor of 12, leading to coefficients comparable with Dekker's. A suitable method for determining temperature coefficients is described in 2.8.

5.5 Sources of undersized signals

5.5.1 Inelastic scattering

The effect of scattering on the shape of an SER is investigated following the statistical treatment of Coates (1973). Consider the energy spectrum of secondary electrons produced by primary electrons, of energy E_p electron volts, incident on the first dynode. The spectrum shown in Fig. 5.8 would be obtained by repeated measurement of the outcome of a large number of primaries, acting one at a time. To the author's knowledge this has not been done for the highly active alkali materials used in PMTs and we have therefore to assume that the energy spectrum in this case will follow the same general form as those published in the scientific literature for metals, semiconductors, and insulators. A typical output spectrum always includes signals that have suffered a loss from inelastically scattered events early in their formation (Region B in Fig. 5.8). This part of the distribution is not flat as depicted, but is slightly dished. However, ignoring this feature by assuming a flat curve is not considered critical to the treatment that follows.

Assume that a fraction k of primary electrons produce inelastically scattered electrons; this is represented by the area enclosed by the dashed box in Fig. 5.8 and then divided by the total area of the spectrum. According to this basic model, any fraction k of the incident energy between 0 and 1 may be lost to inelastic scattering, and all values of k are equally likely. The energy available to create true secondaries is $(1 - k)E_p$. For fixed E_p , we would expect the measured electron number distribution, in the absence of inelastic scattering, to be a Poisson one with mean $\mu = E_p/e$. The number spectrum, m , allowing for inelastic scattering, consists of the superposition of two distributions: one for those primaries that have suffered inelastic scattering, and another for those that have produced only true secondaries. These have not suffered scattering and are thus assumed to be Poisson distributed. The probability distribution for events in which inelastic scattering has occurred is a Poisson distribution with a reduced mean of $\mu(1 - k)$, given by (5.8):

$$p(m, \mu(1 - k)) = e^{-\mu(1-k)} (\mu(1 - k))^m / m!. \quad (5.8)$$

Integrating over all k , we obtain

$$\int_0^1 p(m, \mu(1 - k)) dk = \int_0^1 e^{-\mu(1-k)} (\mu(1 - k))^m / m! dk.$$

For $m = 0$, for example,

$$\int_0^1 p(0, \mu(1 - k)) dk = \int_0^1 e^{-\mu(1-k)} dk = (1 - e^\mu) / \mu. \quad (5.9)$$

Calling these probabilities $p_k(m, \mu)$, we have, for the first four, after which the pattern becomes obvious,

$$\begin{aligned} m = 0, \quad p_k(m, \mu) &= (1 - e^{-\mu})/\mu; \\ m = 1, \quad p_k(m, \mu) &= (1 - e^{-\mu})/\mu - e^{-\mu}; \\ m = 2, \quad p_k(m, \mu) &= (1 - e^{-\mu})/\mu - e^{-\mu} - e^{-\mu}\mu/2!; \\ m = 3, \quad p_k(m, \mu) &= (1 - e^{-\mu})/\mu - e^{-\mu} - e^{-\mu}\mu/2! - e^{-\mu}\mu^2/3. \end{aligned} \quad (5.10)$$

If a proportion k of the primaries suffer inelastic scattering, then a fraction, $(1 - k)$, are unaffected; the overall probability distribution is given by

$$P_m(\mu) = kp_k(m, \mu) + (1 - k)p_{1-k}(m, \mu). \quad (5.11)$$

Figure 5.13 shows that $k = 0.3$ is a reasonable value for a PMT, and this has been assumed in Fig. 5.14. The distribution for $P_m(\mu)$ with $\mu = 10$, and $k = 0.3$, showing its composite nature, is plotted in Fig. 5.14. The general shape of the predicted SER, in the region 0.2 to 0.5 photoelectrons equivalent, accords with measured distributions for quality photon counting PMTs, as verified for a real PMT in Fig. 5.15. Also, we show in Fig. 5.15 that the distribution beyond the peak closely follows a Poisson distribution, with mean value of $\mu_1 = 20$. Note there are inter-dynode scattering losses but, as cascade theory shows, the added dispersion is minimal.

The distributions in Fig. 5.14 are theoretical. But the shape of a recorded SER may be described as a combination of a degraded Poisson distribution and a scattered component given by (5.11). For this purpose we need to know μ_1 , which may be deduced by the method described in Fig. 5.16, together with an assumed coefficient for k . Figure 5.15 illustrates this for the stated parameter values.

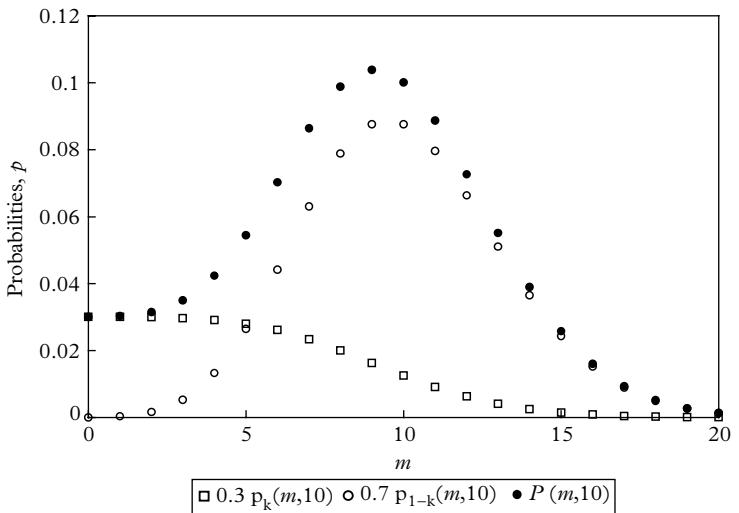


Fig. 5.14. The statistics at the first dynode based on the Poisson distribution, of mean $\mu_1 = 10$, after allowing for inelastic scattering. The composite distributions can be mixed in any proportion of $k:(1 - k)$; in this example, the ratios are 0.3:0.7. The $P:V$ ratio for $P(m, 10)$ is 3.7.

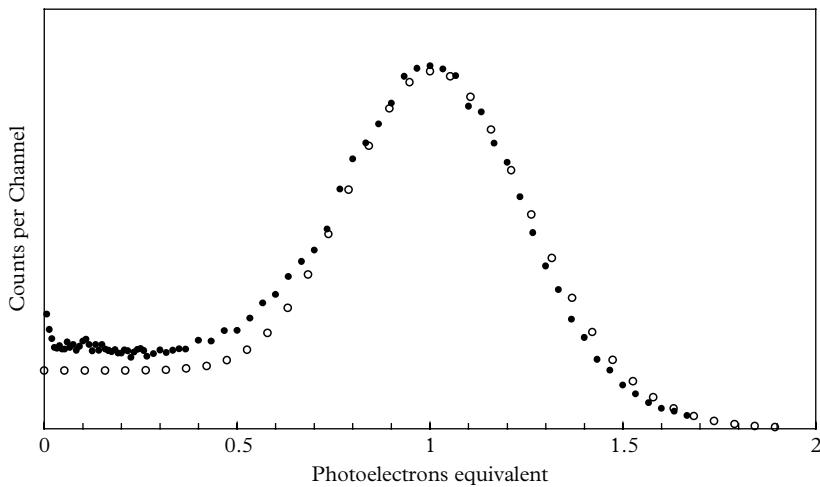


Fig. 5.15. The distribution indicated by the solid dots refers to a photon-counting quality PMT operated with $\nu(k-d_1) = 300$ V, for which $\mu_1 = 20$, and k is taken as 0.24 to attain the fit illustrated. The light source was set off-axis to prevent photoelectron emission at d_1 (see 5.5.3). The predicted shape of the SER (open circles) is remarkably close to the measured one.

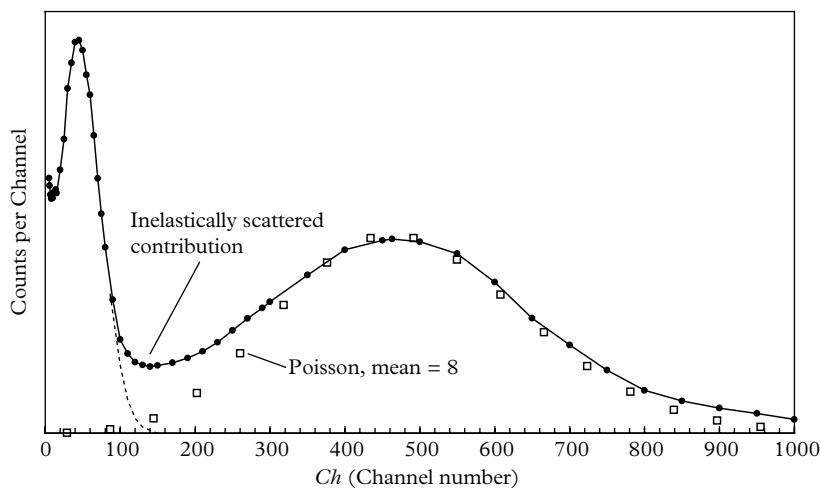


Fig. 5.16. A single-electron response curve for a linear focussed PMT showing a subpeak at Channel 45 attributed to photoelectrons originating from d_1 , and a second peak at Channel 460 for events originating from the photocathode. The accelerating potential $\nu(k-d_1)$ is 100 V, to best accommodate both distributions. The ratio of the two peaks gives the most probable d_1 gain of 10.2. This is close to the mean gain deduced by computation.

5.5.2 Gain non-uniformity

In an attempt to explain these undesirable features of the SER, Baldwin and Friedman (1965) proposed that certain dynode materials may exhibit a heterogeneous secondary emission coefficient over their active surface. In particular, Ag-MgO may be polycrystalline with random orientations whereas caesiated dynodes, common in PMTs, are likely to have more uniform emission properties. They assumed that each microscopic element of the dynode surface may be characterized by a particular mean secondary emission coefficient, with a distribution $f(x) = \exp(-x/\epsilon)/\epsilon$ over the surface. Combining this distribution with a Poisson one leads to a geometric distribution. Prescott (1966) developed this idea using the Pólya distribution given in (5.12), which, because of the parameter, b , has the property of converting to a geometric distribution at one extreme (when $b = 1$) and to the Poisson at the other (when $b = 0$); intermediate values of b generate a composite distribution. To quote Prescott (1966):

the Pólya statistical model describes secondary emission from a dynode of finite area for which the average number of secondaries per incident electron for the first dynode as a whole is μ_1 . For any one point on the dynode the probability distribution of the secondaries is Poisson but the local value of the mean number of secondaries varies from place to place in a manner described by the Laplace distribution [represented by $g(m)$ in Prescott's paper].

The assumption that the secondary emission coefficient varies over the surface of a dynode is a reasonable one, but the problem is that there is no easy way in which to determine this function. There is, however, experimental evidence for an 'effective' spatial variation in secondary emission in the dynodes of a Venetian blind PMT. The term effective is appropriate here because the variation in secondary emission is actually a geometrical one related to collection from the first dynode, and to skipping (see Figs 5.19 to 5.21). Prescott makes this point, stating that the secondary emission coefficient varies, either because the mean number of electrons actually ejected varies from place to place, or the emitted electrons do not have the same probability of reaching the next dynode.

The Pólya distribution for the first dynode, predicting the probability of generating n electrons, when the mean number is μ_1 , is

$$P_b(n) = \mu_1^n / n! (1 + b\mu_1)^{-n-1/b} \prod_{i=0}^{n-1} (1 + i b), \quad (5.12)$$

where b is a parameter, $0 \leq b \leq 1$, and n is an integer > 0 . A family of curves representing (5.12) is shown in Fig. 5.17. The generating function, given by Prescott, is

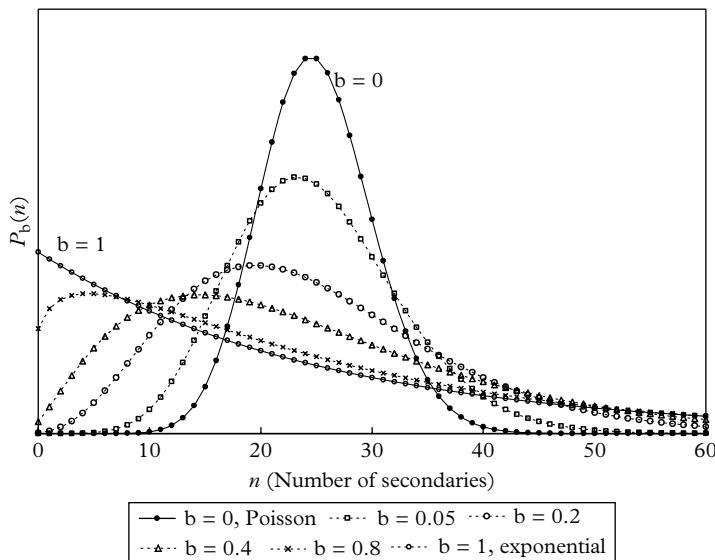


Fig. 5.17. The Pólya distribution for a range of b values between 0 and 1. The mean, μ_1 , is 25 for all distributions. Note that only integers are permitted for n , and the superimposed curves are used to aid presentation.

$$G(u) = \left(1 + b\mu_1(1 - u)\right)^{-1/b},$$

from which the probability of producing zero secondaries, $P_b(0)$, is

$$P_b(0) = G(0) = (1 + b\mu_1)^{-1/b}.$$

The behaviour of this function is of particular interest at the extreme values of b . In (5.12), when $b \rightarrow 0$, the \prod factor is unity and $(1 + b\mu_1)^{-1/b}$ becomes $e^{-\mu_1}$, making $P_0(n)$ Poisson distributed; thus,

$$P_0(n) = \mu_1^n e^{-\mu_1} / n!.$$

At the other extreme, with $b = 1$, the \prod factor is identical to $n!$, and

$$P_1(n) = \left(\mu_1 / (1 + \mu_1)\right)^n / (1 + \mu_1).$$

This is a geometric series, and we can show using the approximation $\ln((1 + \mu_1)/\mu_1) = 1/(1 + \mu_1)$ that it approximates a negative exponential distribution,

$$P_1(n) \approx 1/(1 + \mu_1) \exp\left(-n/(1 + \mu_1)\right), \quad (5.13)$$

provided that μ_1 is greater than about 10. Equation (5.13) provides the justification for the observation that the Pólya distribution includes as limiting cases a

Poisson at one extreme ($b = 0$) and an exponential distribution at the other ($b = 1$). The author's view is that attempts at fitting the Pólya distribution to measured SERs of the type shown in Figs 5.5 to 5.7 are unconvincing, particularly for small b . The major source of small pulses lies in backscattering rather than inhomogeneous stage gain.

5.5.3 Events initiated by d_1

A sharp rise in the distributions at very low pulse heights is evident in Figs 5.5, 5.6, and 5.7. There may be a significant noise contribution from a preamplifier, if included in the detector, particularly if it is operated at high sensitivity. This is essentially an 'electronics problem' for which the solution lies in reallocating the overall gain, as discussed in 14.1. Photoelectrons are also produced by light transmitted through the photocathode and terminating on the first dynode. This contribution is easily verified by aiming a thin pencil of light on to the first dynode, to produce the results shown in Fig. 5.16. The counts in the low-energy peak have been maximized by using red light incident axially. Further confirmation of the origin of the subpeak follows by open circuiting the cathode, in which case the main peak disappears and the subpeak distribution remains and follows the dashed curve shown. The subpeak largely disappears on launching the light off-axis. The ratio of the peak channel numbers is a measure of the most probable gain of the first dynode and, since the two distributions have similar shapes, the mean d_1 gain is also given approximately by this ratio. The contribution to undersized pulses depends on the wavelength of light and on the ratio of the active area of the first dynode to that of the photocathode. The latter is of the order of 0.25 for a standard 1" PMT but only ~ 0.01 for a 5" one. These d_1 -initiated events are generally of minor concern in spectroscopy but they appear as prepulses in timing apparatus, where they are significant, although at a low level of less than 1% of the mean rate (see Chapter 11, Fig. 11.4).

5.5.4 Other sources of undersized signals

The acceleration experienced by an electron moving in the vicinity of a nucleus, of atomic number Z , creates radiation known as bremsstrahlung, the intensity of which varies as Z^2 . The incident electron can radiate any amount of energy from zero up to its total kinetic energy. The radiated energy spectrum is continuous with superimposed X-ray lines, for example, the K and L fluorescence peaks; these result from photoemission following direct impact of the incident electron with a bound one. Specifically, when a primary electron emits bremsstrahlung within the surface of the first dynode, the latter acts like the anode of a conventional X-ray tube and, analogously, the yield is inefficient. The energy loss by electrons from low-energy X-ray production, at energies appropriate to PMTs (< 1000 eV), is less than 0.1% of that from ionization and excitation (Evans 1955). The

generation of a soft X-ray, which subsequently escapes into the vacuum, will lead to an undersized event in the same manner as an inelastically scattered electron does, but this contribution to the low pulse height region of the SER is negligible.

The ‘edge effect’ has been cited by Coates (1970a), and Oliver and Pike (1970), to account for some of the low pulse heights characteristic of all SERs. Quite simply, a fraction of the secondaries from d_1 , because of their unfavourable initial energy and emission angle, will miss the active part of d_2 and land on the ceramic support structure; others may skip d_2 and land on some other dynode, such as d_3 . The edge effect mechanisms mentioned here show up in electron-optical simulations, but the proportion of events so affected remains unknown. Evidence of the edge effect can sometimes be inferred from studying the change in shape of the SER as a function of the position of illumination on the photocathode, as demonstrated by Oliver and Pike (1970).

5.6 Discrete dynodes

With the exception of the Venetian blind structure, the first dynode of a discrete multiplier stack usually has a larger surface area than the others, together with a different orientation from the common dynodes. In a Venetian blind structure, all dynodes, except for the last, are identical in geometry. The first dynode, in all multiplier types, must provide efficient collection of photoelectrons, while directing secondary electrons to the next dynode, ideally without loss. The Venetian blind dynode has a large aperture and it is thus able to collect photoelectrons with high efficiency—particularly relevant in large-diameter PMTs. However, its asymmetry in the form of slats is a distinct disadvantage, as will be discussed. The last dynode is different from all others: it consists of a flat plate mounted in proximity to the anode. The so-called reflex anode consists of a fine mesh, located between the penultimate dynode and the last dynode. Electrons from the penultimate dynode pass through this mesh on their way to the last dynode. The mesh is located close to this dynode to create a strong extraction field, which benefits linearity and speed of response. Finally, the first dynode tends to have higher gain than the common ones. For the analyses of this chapter, it is sufficient to assume that all dynodes in a particular multiplier behave in the same way.

There is a range of PMTs for which gain is derived from a continuous dynode system. Channeltrons™ utilize a resistive secondary emitter coated on the inside surface of a hollow glass tube, the length of which may extend up to several

centimetres. Multichannel plates consist of the fusion of many microscopic elements into a matrix containing up to one million channels—described in greater detail in 5.7.2. While continuous dynode multipliers have rather different properties from the discrete types, their amplifying operation is based on secondary emission.

5.6.1 Venetian blind dynodes and skipping

A poorly defined peak in the SER may be expected where low-gain dynodes are used. In particular, electron multipliers with either BeCu or AgMg dynodes require 15 to 20 elements to provide gain comparable with that of a typical 10-stage PMT, with SbCs dynodes. Venetian blind multipliers are more likely to exhibit quasi-exponential distributions than other dynode types. The explanation for this lies in the electron-optics of the dynodes. The equipotential lines within the slats of a Venetian blind dynode, shown in Fig. 5.18, reveal a weak extraction field gradient at the top of the slats. This is of the order of only 1 V/mm that progressively strengthens as the lower edge of the slat is approached. Only those secondary electrons with a fortuitous combination of emission angle and launch velocity escape to the next dynode, as illustrated in Fig. 5.19; the effective secondary emission coefficient is therefore a function of the landing position of the incoming electron on the slat.

The inability to collect uniformly over the entire surface area of a dynode slat explains why secondary emission from Venetian blind dynodes has a broader

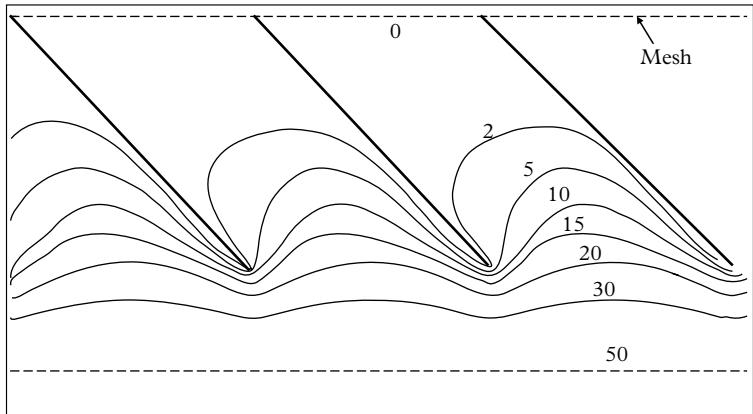


Fig. 5.18. The extraction field in a Venetian blind dynode is provided by the potential of the mesh attached to the dynode below (maintained at 50 V in this example). The upper region of a slat is effectively shielded by the slat itself.

number distribution than would be predicted from purely statistical considerations. This phenomenon has been nicely demonstrated by Kuroda *et al.* (1981), who observed that a modulation in the anode output corresponded to the spatial periodicity of the dynode slats. The anode output, for electrons released from the top of the d_1 slat, as in Fig. 5.19, is about half the magnitude of the signal corresponding to emission from the bottom tip.

There is a source of broadening caused by dynode skipping, in which a primary electron traverses the space between the slats without interacting with any part of the dynode. This defect is most serious in the first dynode, as illustrated in Fig. 5.20. Secondary electrons may skip any of the dynodes in the stack (except for the last one), but clearly the most serious consequences stem from the first dynode. The cause lies in the asymmetry of the region between the photocathode and first dynode, depicted by the trajectories marked (a) and (b) in Fig. 5.20. The magnitude of the effect can be quantified by using a flying spot, 1 to 2 mm in diameter. The trace in Fig. 5.21 corresponds to a traversal across a diameter of the photocathode, following a path at right angles to the direction of the slats. The loss in anode output, corresponding to Region (a) of the trace, is entirely due to skipping.

An alternative way to quantify dynode gain non-uniformity is to study the variation in the SER as a function of the point of illumination of the photocathode,

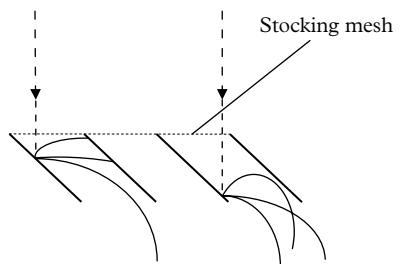


Fig. 5.19. The efficiency of extraction from a Venetian blind slat depends on the landing point of the incident electron, shown by the dashed line. Secondary electrons are extracted by the field created by the mesh of the dynode below the one illustrated, but those secondaries with unfavourable launch angles, or high initial energy, land on the back of the neighbouring slat and fail to escape. The secondary emission coefficient is highest for primary electrons that strike the bottom of the slat.

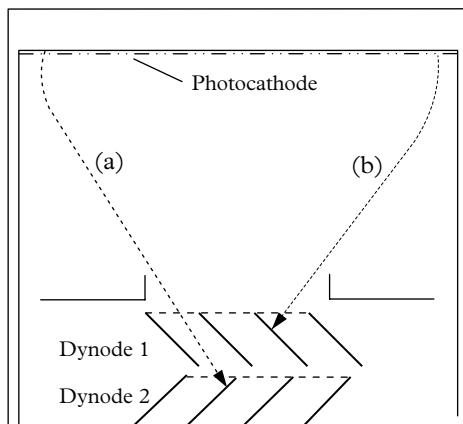


Fig. 5.20. Dynode skipping poses serious amplification non-uniformity in Venetian blind dynode structures. The output signal initiated by a photoelectron following trajectory (a) will be undersized compared with the one which follows trajectory (b). Photoelectrons with trajectories approximately parallel to a slat have a high probability of skipping that slat, but those originating from a point diametrically opposite (b) will always hit a slat of the first dynode.

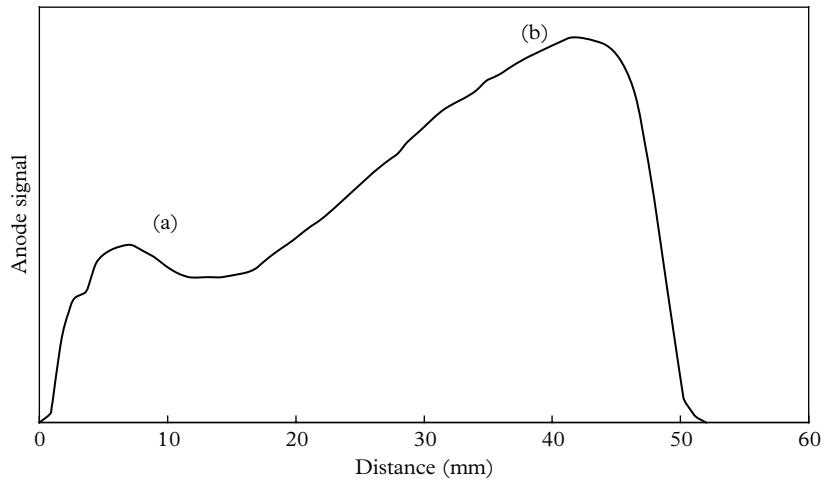


Fig. 5.21. Illustrating the positional non-uniformity for a 2" Venetian blind PMT. The size of the anode signal is influenced by the position of illumination on the photocathode. The signal at (a) is about half of that at (b) because of skipping when the flying spot scan is at right angles to the directions of the slats; (a) and (b) correspond to the trajectories of Fig. 5.20.

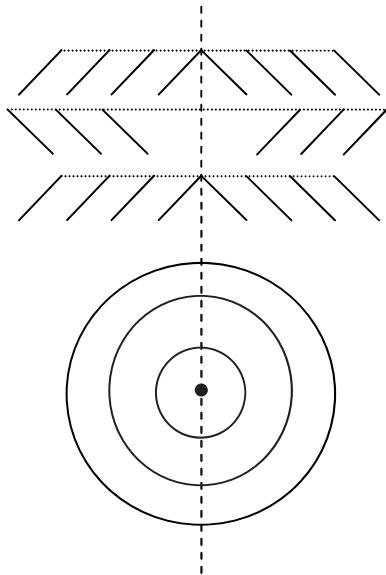


Fig. 5.22. Two views of the target dynode structure show how the asymmetry inherent in the Venetian blind dynode may be eliminated.

shown in Fig. 5.7. There is useful information in the variation in the total count rate, and in the shape of the SER. Given a point light source of single photons of constant output, the variation in count rate as a function of point of illumination on the photocathode with this light source tells us something about the collection efficiency. Further discussion of this topic is reserved for 10.3, but for the present we are concerned only with the shape of the SER. The part of the trace marked (a) in Fig. 5.21 corresponds to the poorest of the SERs in Fig. 5.7. It is difficult to establish the mean gain for this SER with any certainty, because of the predominance of small pulses in the spectrum, but it is clear from inspection that it must be only about half of that for the peaked distributions that are included. The SERs for other points of illumination are reasonable, but the distribution for full flood exhibits a shoulder rather than a distinct peak. The set of curves in Fig. 5.7 explains why some authors do not record a peak in the SER because of the dependence on the area of the cathode actually illuminated.

The phenomenon of dynode skipping was proposed by Delaney and Walton (1964) in their work on the shape of single-electron spectra, and the results taken by the present author in Figs 5.19 to 5.21 fully support their conclusions. The asymmetry inherent in the conventional Venetian blind PMTs can be corrected by adding circular symmetry to the dynode structure, illustrated in Fig. 5.22. This

was patented by SRC Laboratories Inc. (British Patent Number 1 556 516, Nov. 1979) and used in their 2" and 3" PMTs for gamma cameras. Although the patent has long since expired, no manufacturer currently uses this structure, perhaps because of difficulty in fabrication together with a declining market for Venetian blind PMTs.

5.6.2 Coarse mesh dynodes

Kuroda *et al.* (1981) developed a position-sensitive multiplier using a standard 2" Venetian blind PMT and an axial magnetic field. The basis for this approach was the realization that the envelope of the secondary electron cascade, created by each photoelectron, substantially maintains its (x_0, y_0) coordinate as z , the cascade depth, increases towards the anode. The slats in this invention are triangular in cross section, 1 mm in width, and separated by a similar distance. Each stage consists of two BeCu meshes offset from one another, also by 1 mm, as shown in Fig. 5.23. Field lines and electron trajectories are shown in Fig. 5.24, illustrating how position sensitivity is maintained and it is clear from comparing the field patterns in Fig. 5.24 with those in Fig. 5.18 that the extraction field for triangular slats, with the more open structure, is far superior to that for the conventional Venetian blind structure. The role of the magnetic field is to channel the paths of the electrons in a tight spiral around the magnetic field lines, which are parallel to the z-axis.

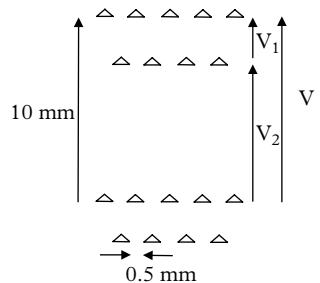


Fig. 5.23. The pitch in these coarse mesh dynodes is 10mm with an inter-dynode voltage of typically 150 V. The gain per stage is not very sensitive to the disposition of V_1 and V_2 and typically $V_1 = V_2 = V/2$.

Kuroda, K., Sillou, D., and Takeuchi, F. (1981). New type of position sensitive photomultiplier. *Rev. Sci. Instrum.*, 52(3), 337–346. Reprinted with permission from AIP© (1981).

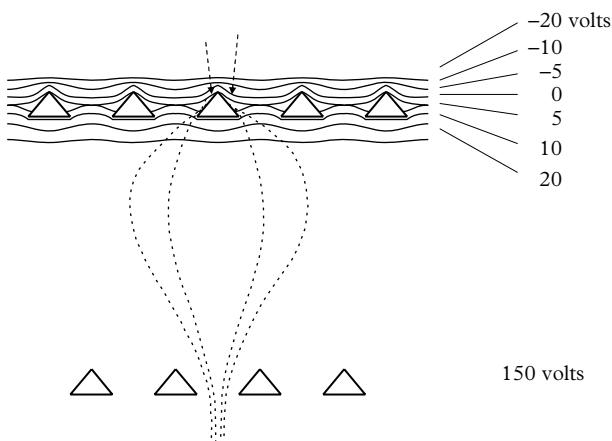


Fig. 5.24. Detail of part of the dual mesh configuration that constitutes each dynode, showing electrostatic field patterns and focussing of trajectories.

Agoritsas, V., Kuroda, K. and Nemoz, Ch. (1989). A new type of position-sensitive electron multiplier. *Nucl. Instr. and Meth. in Phys. Res. A*, 277, 237–241. Order number 3620250895810. Reprinted from Elsevier Science ©1989.

Further exploitation of this dynode type was reported by Agoritsas *et al.* (1989), who constructed a 14-stage multiplier for the detection of ions using bar widths and spacing down to 0.2 mm. The multiplier of Kuroda *et al.* (1981) was incorporated into a bialkali PMT but, unlike the traditional Venetian blind PMT, the speed of response for the coarse mesh type is considerably faster, with a rise time of 1.8 ns. Both devices include multi-anode strips providing spatial resolution of the order of a few millimetres, without an applied magnetic field. The Hamamatsu position-sensitive PMTs, R2486-02 and R3292-02, incorporate dynodes of similar construction.

5.6.3 Fine-mesh dynodes

PMTs with this type of dynode were introduced in the 1980s to meet the demands of certain high-energy physics applications. The gain requirements were modest and could be met by three- or four-stage multipliers, referred to as triodes and tetrodes, respectively. The challenge of operating in magnetic fields of the order of a tesla called for close proximity dynodes and hence the use of fine meshes, depicted in Fig. 5.25. Neighbouring meshes are offset to prevent light and ion feedback but skipping is clearly of significance, especially in the first dynode, and a non-peaked SER is likely. Higher-gain, 19-stage PMTs are available with gain capability of up to 10^7 in a field of ~ 0.5 T. Inclusion of cross-wire anodes gives positional sensitivity.

5.6.4 Metal-channel dynodes

Metal-channel dynodes, shown in Fig. 5.26, are incorporated in Hamamatsu R5600 series of metal-can, compact PMTs, first reported by Kyushima (1993). The range of PMTs that include this structure is impressive, with the addition of the multi-anode (4×4) type R7600-03-M16 to the family. The SER shown in Fig. 5.27 has been reported on by Abbon *et al.* (2008), showing two distinct peaks

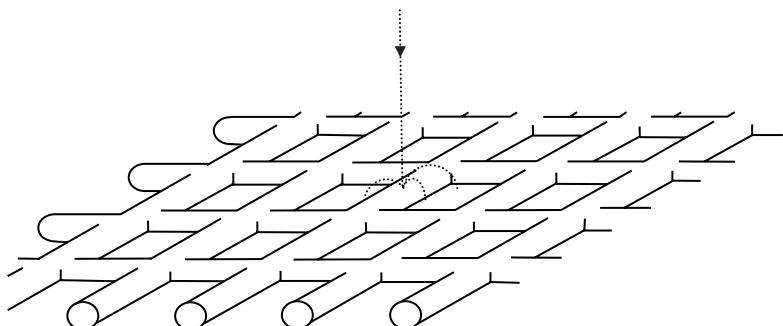


Fig. 5.25. A fine mesh leads to compact designs with a dynode pitch of the order of 1 mm, with fast response, position sensitivity, and good pulse height linearity.

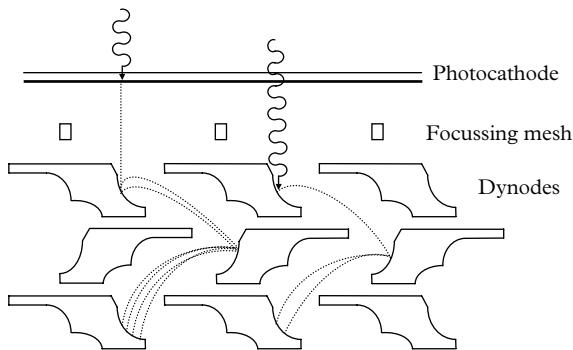


Fig. 5.26. The diagram shows a cross section of three metal-channel dynodes used in the TO8 range of compact PMTs. Photons transmitted through the window may generate a photoelectron from d_1 ; the effect is quite pronounced at long wavelengths, beyond 550 nm. If the SER is measured with light of a different wavelength from that in the application, clearly the SER will not be representative at other wavelengths. Skipping occurs but only for off-axis photoelectrons. The purpose of the focussing mesh is to concentrate the photoelectrons on the active part of d_1 .

Chirikov-Zorin, I., Fedorko, I., Menzione, A., Pikna, M., Sýkora, I., and Tokár, S. (2001). Method for precise analysis of the metal package photomultiplier single photoelectron spectra. *Nucl. Instr. and Meth. in Phys. Res. A*, 456, 310–324. Reprinted from Elsevier Science ©2001.

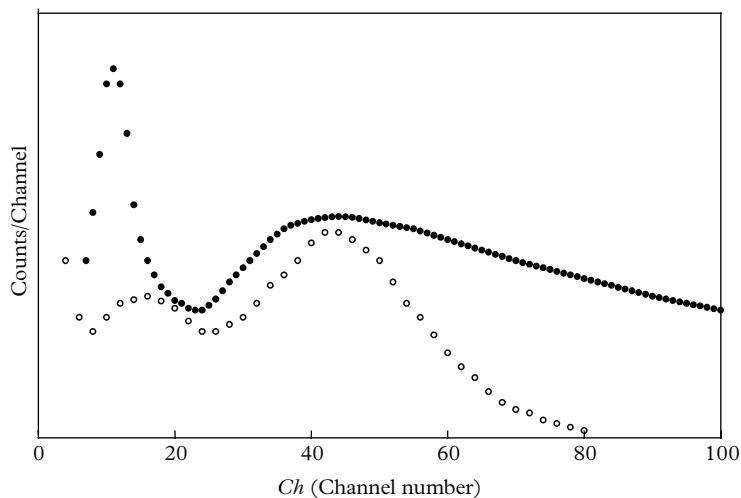


Fig. 5.27. SERs measured by Abbon *et al.* (2008; ○), using light of wavelength 360 nm, and by Chirikov-Zorin *et al.* (2001; •), at ~620 nm. The broad distribution of the upper SER (•) is consistent with a mean d_1 gain of ~3.5, as indicated by the ratio of the peak to the subpeak. The SER of the lower distribution (○) is considerably sharper, implying a d_1 gain of about 10.

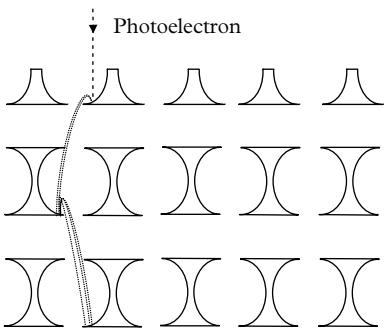


Fig. 5.28. The Photonis foil dynode system consists of thin perforated foils arranged in parallel sheets.

for single-photon excitation; the lower energy subpeak occurs at one-third of the position of the main peak, and it is attributed to skipping by the authors. A second SER shown in Fig. 5.27, also exhibiting a subpeak, is reported by Chirikov-Zorin *et al.* (2001), who used a red LED in their investigations; they ascribe a subpeak to light transmitted to a photosensitive d_1 . This explanation is consistent with the present author's measurements and conclusions regarding Fig. 5.16. There is also a contribution to the subpeaks from backscatter.

One of the outstanding features of PMTs made with these dynodes is their compactness derived from a dynode pitch of just over 1 mm. Photonis offered a position-sensitive PMT, with so-called foil dynodes, shown in Fig. 5.28, for use in high magnetic fields. The dynode profiles are different from the metal-channel type, but the concept is essentially the same.

5.6.5 Linear and circular focussed dynodes

V J Zworykin and J A Rajchman (1939), RCA staff members at that time, described two new multipliers in a paper headed 'The electrostatic electron multiplier'. It is remarkable that these dynodes, known as linear focussed and circular focussed, are still used in essentially their original form in currently available PMTs. Equally remarkable is that these authors worked in the pre-computer age, making use of an electrolytic tank, a stretched rubber membrane, and various graphical means to determine the electrostatic field lines and the corresponding electron trajectories. Three-dimensional electron-optics programs are now available from several sources for simulating the paths of electrons, with launch angles, initial energy, and predicted transit times included. These simulations are readily extended to positive ions for afterpulse investigations, for example.

The cross section of a common linear focussed dynode is an arc of a circle extended on one edge by a tangential segment; the two curved ends are enclosed, making each dynode a shallow bucket, as shown in Fig. 5.29. This prevents secondary electrons from spreading towards the ceramic side plates. The continued pre-eminence of these two dynode types is assured by two outstanding performance features: dynamic range and isochronous electron transit times. Strong extraction field gradients at the surface of each dynode ensure that the onset of non-linearity, because of space charge saturation, occurs at considerably higher pulsed currents than is the case for box-and-grid dynodes, for example. With linear focussed dynodes, there is a compensating effect in which a longer electron path between, say, d_{n-1} and d_n is balanced by a shorter path between the next pair of dynodes, d_n and d_{n+1} . This can be understood in terms of the crossing-over of trajectories evident in Fig. 5.30.

The configuration shown in Fig. 5.31 evolved shortly after the original design of 1939, yet contemporary side window designs are little different. Conceptually, the circular focussed dynodes are the same as linear focussed ones, except curled up for compactness. These features of the design, together with the transit time compensating effect, create a fast dynode stack.

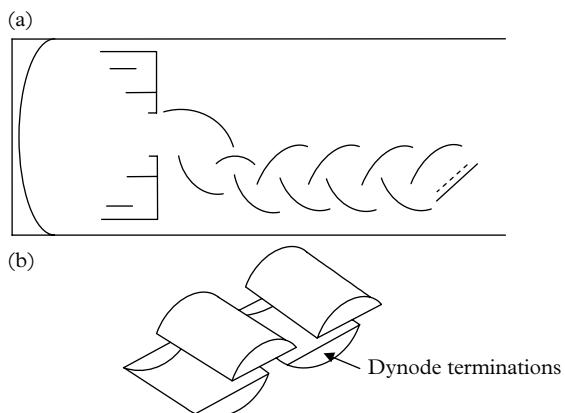


Fig. 5.29. (a) Showing the cross section of a linear focussed dynode system incorporated in fast PMTs where a typical 12-stage PMT has eight common dynodes, depicted in (b). The first dynode, with a larger area than a common dynode, is designed to optimize photoelectron collection; the second and third dynodes serve to couple d_1 to the start of the common dynode set. The last dynode is a plate located in close proximity to the anode mesh—its purpose is to preserve pulse height linearity and speed of response.

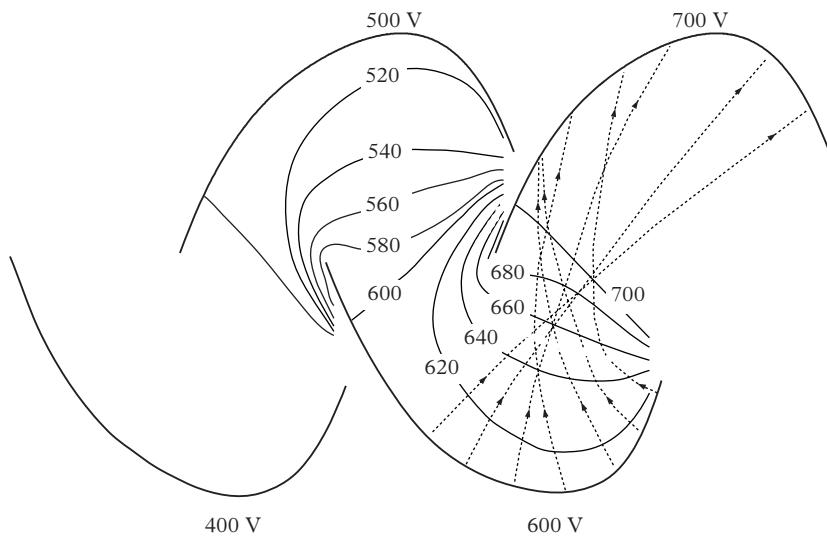
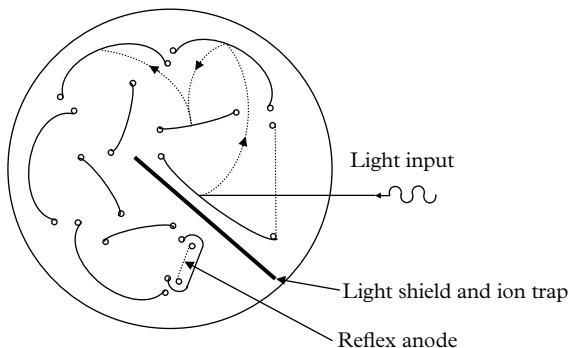


Fig. 5.30. Illustrating the strong extraction field provided by linear focussed dynodes. The thin lines are equipotentials from which we deduce an extraction field of 20 V/mm; compare this to the figure of 1 V/mm for a box-and-grid dynode. Electron trajectories, shown by the dashed curves, cross over, thereby compensating the overall transit time of the PMT. The inter-dynode transit time is of the order of 2 ns for the linear and circular focussed types.



5 Fig. 5.31. Circular focussed dynodes were incorporated in the first commercial PMTs, the 931A and the IP28. Note the photocathode is the first electrode of the structure. A hairpin anode was adopted in the earlier designs.

5.6.6 Box-and-grid dynodes

The common element in this dynode array is a quadrant-shaped box, shown in Fig. 5.32, the entrance aperture of which is covered by a fine mesh while the exit is open. A PMT that includes this type of dynode exhibits a well-defined SER but the pulse rise time of 15 ns is long compared with the performance of other dynode types. Electron-optics simulations in Fig. 5.33 indicate a weak extraction field over the entire dynode surface: the electric field gradient is 1 to 2 V/mm, which is about one-tenth that of a linear focussed dynode. This limiting performance occurs despite the inclusion of meshes, and their omission causes a serious reduction in field gradients.

5.6.7 Transmission dynodes

Dynode structures so far discussed operate in a reflection mode: the secondary electrons are emitted from the surface upon which the primaries are incident. A transmission dynode consists of a thin film ranging in thickness between 1 and 10 μm . Ionization is produced along the entire path of each primary, but only those electrons that diffuse to the surface may escape under the influence of an applied electric field. These electrons are emitted from the opposite face to the one presented to the primary; hence, it is clear that suitable materials are those with a relatively long diffusion length, or equivalently mean free path, to allow sufficient electrons to diffuse to the surface. Purely mechanical considerations favour a thick film for achieving a sufficiently robust structure but this inevitably results in a low yield of electrons.

Gutierrez *et al.* (1972) investigated thin epitaxially grown p-doped GaAs films of thickness 3 to 5 μm , activated for negative electron affinity (that is, caesiated).

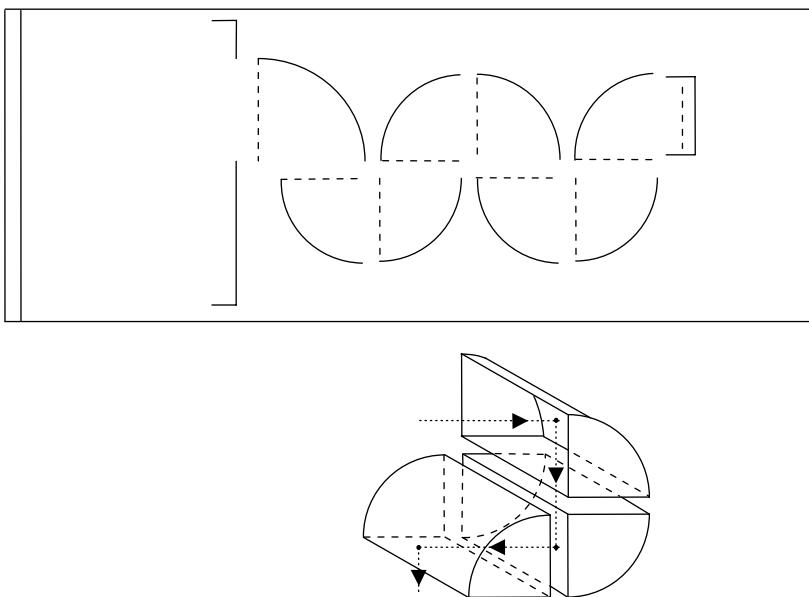


Fig. 5.32. Box-and-grid structure of a nine-stage PMT. In certain PMTs, the first box is bigger than the remainder to facilitate collection from the photocathode. The leading aperture of each box includes a honeycomb mesh to extract the electrons from the previous dynode.

5

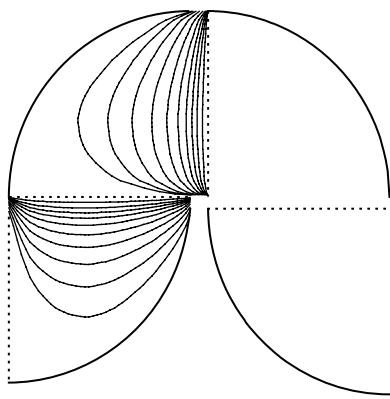


Fig. 5.33. Equipotentials in the box-and-grid structure. The field lines are shown in steps of 10 V for an inter-dynode voltage of 100 V. The speed of response and the linearity are inferior compared with those in other dynode systems.

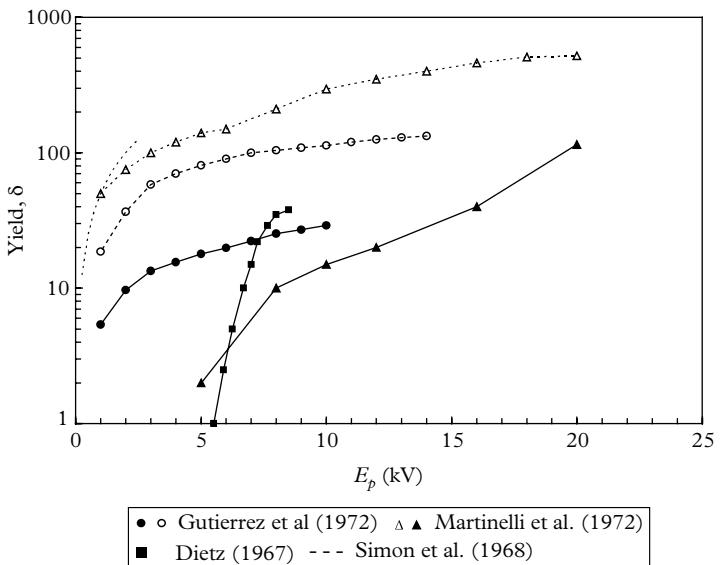


Fig. 5.34. Yield curves for transmission dynodes are shown by the solid lines, while the dashed curves refer to reflection. Martinelli *et al.* (1972) and Gutierrez *et al.* (1972) made both reflected and transmission measurements on their respective samples. Note, for two of the transmission curves, the yield exceeds unity only for $E_p > \sim 5$ kV.

This type of semiconductor is characterized by long diffusion lengths, L, in the range of several microns, compared with escape depths of only 20 nm for metals and insulators (Simon and Williams 1968). The results of Gutierrez *et al.* (1972) for reflection yield are similar in magnitude to those of Simon and Williams (1968) for GaP, but the yield in transmission mode is a factor of 5 lower at approximately 20 at 6 kV. The attainment of high electron yields from these materials is shown in Fig. 5.34. Also shown are transmission measurements by Martinelli *et al.* (1972), who used films of a similar thickness to those of Gutierrez *et al.* (1972), attaining a yield of about 10 at 10 kV. These two investigations used just a single dynode, for which the yield is insufficient to allow an SER measurement.

Reynolds (1966) investigated a five-stage image intensifier (described by Slark and Woolgar (1962)) comprising KCl transmission dynodes; these were operated at 7 kV/stage, giving a gain of around 6. Dietz *et al.* (1967) obtain similar gain figures, also for a single stage KCl dynode used in combination with a conventional 14-stage electron multiplier. Reynolds (1966) reports, somewhat despondently, ‘The size distribution of anode pulses from single photoelectrons decreases monotonically but deviates significantly from an exponential law’. This description applies equally to the distribution measured by Dietz *et al.* (1967), for which no peak is evident in the SER.

The attraction of transmission dynodes lies in their use for proximity focussed imaging devices, where in theory they would perform much as a MCP. The III-V dynodes discussed here have a diameter of about 5 mm, which would restrict their widespread application, but no doubt larger-area dynodes could be fabricated. Other disadvantages of GaAs are its light sensitivity, which approaches 1000 $\mu\text{A}/\text{lm}$, and its dark current of $\sim 10^4$ electrons/s cm^{-2} . The backscatter coefficient k for GaAs with a mean Z of 32 is relatively high at 0.3 and, for KCl, k is about half this value, based on Z_{eff} in 5.4.4. One would expect backscattering to play a similar role in determining the shape of the SER in transmission dynodes as it does in the reflection mode: that is, the creation of an excess of undersized events but without obscuring the peak of the SER. The gain attainable in both the KCl devices under discussion is sufficient to give a peaked SER but the slow start in the yield curve of Dietz *et al.* (1967), commencing at $E_p = 5$ kV, suggests that L is much smaller than the thickness of the KCl layer. There is also an energy loss of ~ 2 keV in traversing the aluminium backing layer (required to suppress light feedback), and the combination of two statistical effects is the likely explanation for the observed pulse height distribution. The fundamental problems associated with the use of transmission dynodes account for their absence in commercial devices.

5.7 Continuous multipliers

The channel electron multiplier (CEM), MCP, and discrete electron multipliers generate gain through secondary emission. For continuous channel devices, we can envisage a cascade of electrons progressing down the channel while increasing their number as the anode is approached. The growth of the cascade is fed by earlier interactions of electrons that produce secondaries from the wall of the device. An average δ applies to each interaction but clearly the concept of an equivalent to a multiplier with n dynode stages is questionable. However, the longer the channel, the higher the gain and in this way channel length and n are loosely related. It is found that the gain–voltage characteristics of both discrete dynode multipliers and continuous channel devices can be expressed as a power law of operating voltage. An MCP is a compression of many thousands of single channels, offering a larger detection aperture for primary electrons.

5.7.1 CEMs

A CEM, illustrated in Fig. 5.35, comprises a semiconducting layer of SiO_2 deposited on the inside of a hollow glass tube of fine bore. Invented by the Bendix Corporation in the USA, it was called a Channeltron ®; the business was bought by Burle and subsequently acquired by Photomis. Electron gains of up to about 10^7 are attained by applying 2 to 3 kV between the ends of the channel. The secondary

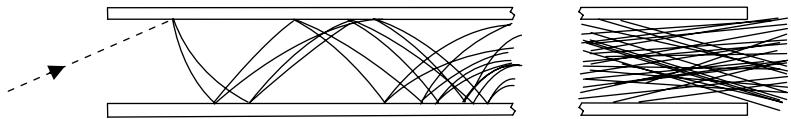


Fig. 5.35. A continuous channel multiplier, showing an electron incident from the left. Multiplication takes place on the walls of the tubing through secondary emission. The wall coating is highly resistive, limiting high count-rate operation.

5

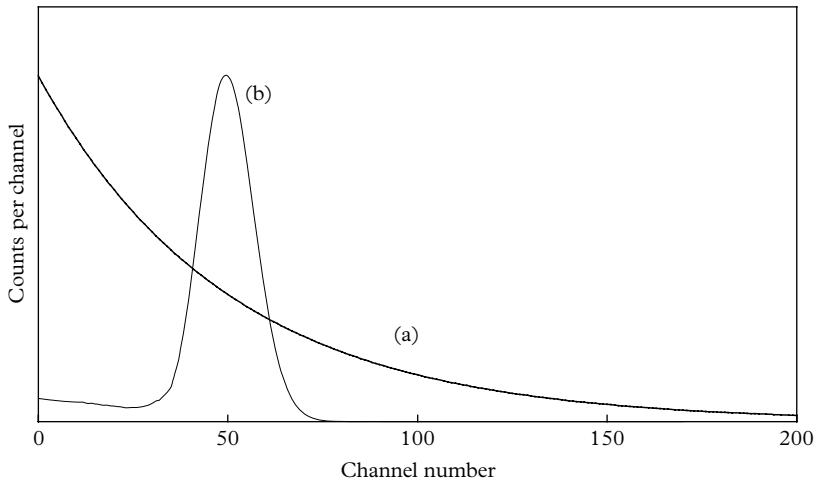


Fig. 5.36. The shape of the SER for CEMs and MCPs depends critically on the mode of operation. (a) Under linear gain conditions the distribution is quasi-exponential. (b) A distinct peak characterizes saturated gain operation.

emission coefficient is only about 2/wall impact, which, together with the variability in the total number of impacts from event to event, means that the multiplication statistics are poor. The SER for a single channel device, or for any one channel of an MCP, resembles a negative exponential under linear gain conditions. This is illustrated in Fig. 5.36(a). A peaked SER (Fig. 5.36(b)) is obtained, by operating in charge saturation mode. At sufficiently high gain, all single-photon-initiated pulses tend to produce the same output charge, leading to a sharply peaked distribution together with a peak-to-valley ratio of 20 or greater. SERs representing the extremes in resolution are given in Fig. 5.36 and further spectra may be found in the literature (Schmidt and Hendee 1966; Klettee *et al.* 1969; Ray and Barnett 1970; Pietri 1975). Note that saturated operation with CEMs is relevant to photon counting and for low-level DC detection applications only.

Single channel devices principally find application in the direct detection of X-rays and ions. The channel is usually curved or serpentine-shaped to frustrate both light and ion feedback, thus permitting high-gain operation. They are used

mainly as ion detectors in mass spectrometers, although one manufacturer, (Perkin Elmer Optoelectronics) offers a small range of PMTs incorporating this type of multiplier. The linearity of performance is limited by the onset of charge saturation with concomitant long recovery time (dead time).

The parallel-plate channel multiplier is a variant of the CEM in which a pair of plates replaces the single channel. The one described by Nilsson *et al.* (1970) consists of two glass plates that are 20 mm × 20 mm, separated by 0.6 mm, and covered on the inside with a secondary emitting layer. SERs with a resolution of 16% at 3.4 kV may be achieved through charge saturation. Several papers have been published on this type of device but none appears to have led to commercial product.

5.7.2 MCPs

The MCP, shown in Fig. 5.37, evolved from the single channel device; it consists of an aggregation of single channels, fused contiguously. An article by Wiza (1979), although dated, provides a useful review of the subject, with clear descriptions of their operation. Individual channels range in diameter from 2 to 40 µm and the plates, with a channel diameter of 2 µm, may contain up to several million elements. Plate thicknesses are typically submillimetre, with a metal electrode evaporated onto each face. The apparent disadvantage of saturation in the CEM devices is exploited in both the two and three stage MCP PMTs to provide an SER

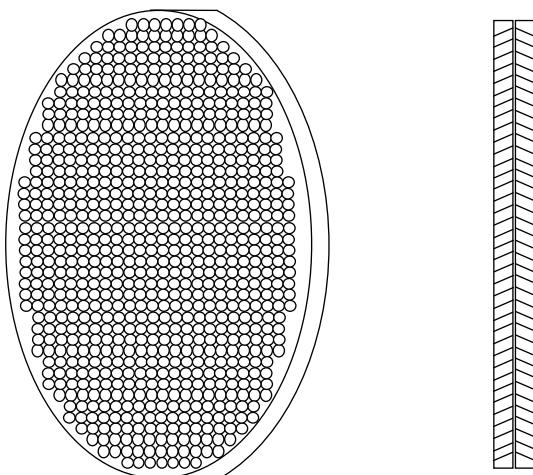


Fig. 5.37. A representation of an MCP showing the individual channels. In practice, the channels are considerably narrower than depicted. They are usually angled, typically at about 10° with respect to the face of the device, to reduce optical and ion feedback. This also leads to higher gain in proximity devices where the photoelectron trajectories are essentially axial to the channels. The two stage chevron arrangement shown on the right offers higher gain than the single stage equivalent (see Fig. 5.38).

of very low variance, and hence multi-photoelectron discrimination. Clearly, should two or more photoelectrons simultaneously enter the same channel, the resulting output pulse height will be indistinguishable from that initiated by a single electron. The occurrence of such events in an MCP with millions of channels is unlikely at low light levels, or where the signal is dispersed over a significant area of the photocathode. But this does not apply when the light is concentrated through focussing and at sufficiently high light levels: the anode output ceases to be linearly related to the intensity of the light input.

As pointed out by Wiza (1979), the performance characteristics of single channels are not a function of channel length l or channel diameter d , but rather of the ratio L/d , and this provides scope for size reduction with the benefits that smaller pore diameters have to offer. First, the adoption of small pores leads to faster time response, with sub-nanosecond rise times easily attainable. Second, spatial resolution, which is important in imaging devices, clearly improves as the pore size is reduced. Third, the dynamic range can be extended by taking advantage of the reduction in both channel capacitance and resistance, leading to a reduced recharging time constant; there is a further benefit since the total number of channels participating per event will consequently increase.

5.7.3 Gain–voltage considerations in MCPs

The gain–voltage characteristic for a single plate 10 μm MCP operated in linear mode is given in Fig. 5.38. The same functional dependence on V , as for discrete dynode multipliers, applies. The straight line is the relationship $g = (1.43 \times 10^4) V^{9.9}$,

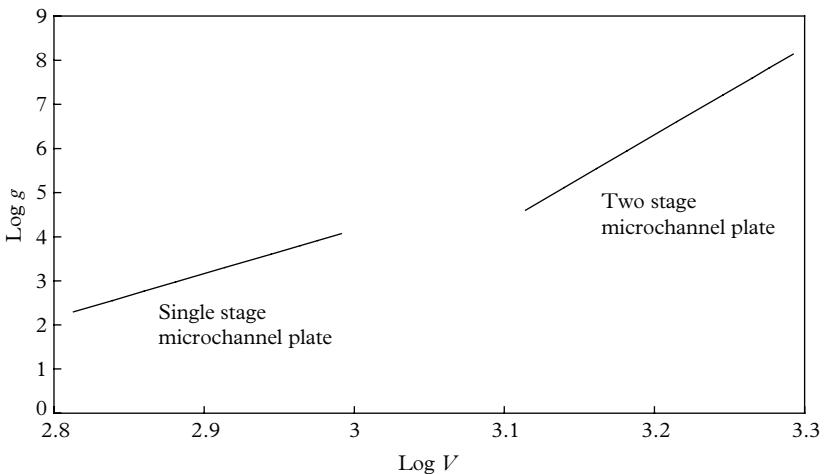


Fig. 5.38. The gain–voltage characteristic of the HPK, F6585 singleplate MCP. The left-hand line is a power-law fit to the manufacturer’s data. The curve on the right gives an indication of the expected characteristic of two such plates in a chevron configuration.

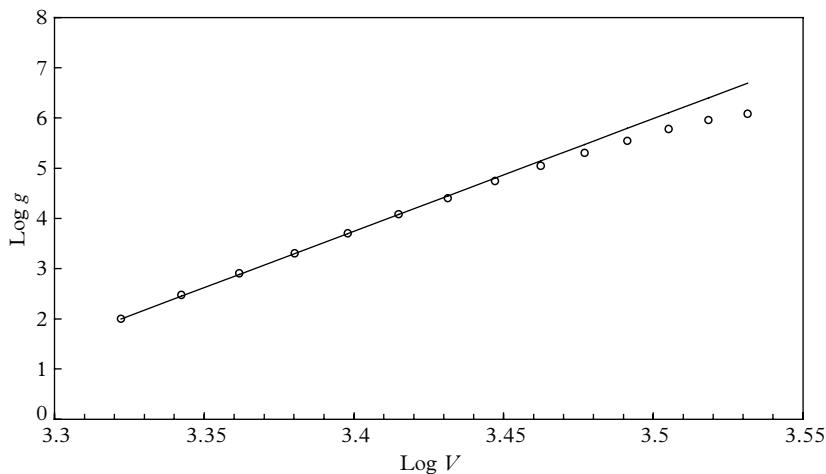


Fig. 5.39. The two stage MCP, type R3809U, shows a flattening of the power relationship at high gain. The open circles are taken from the manufacturer's data sheet.

where V is expressed in kilovolts. It is remarkable that both a continuous multiplier structure and a discrete one can be characterized in the same way, when the physical processes involved in the multiplication processes are so different. The secondary-electron emission coefficient, δ , for the channel material is of the order of 2, and a single stage MCP has insufficient gain capability for photon counting, but two of these in cascade have the characteristic shown in the right-hand side of Fig. 5.38. In this instance, for two identical stages, $g = (1.43 \times 10^4) (V/2)^{9.9}$.²

It is clear that such a double-chevron device would have sufficient gain for low-light-level detection, and in particular photon counting. In practice the two stage and three stage devices do not perform quite as predicted, for two reasons. First, there are undoubtedly interface losses when two stages are coupled and, second, the effect of charge saturation is to flatten the curve at high gain. This is illustrated for an actual two stage device in Fig. 5.39. The relationship between g and V is $g = (6 \times 10^{-6}) V^{22.4}$, with V in kilovolts. A well-resolved SER would be obtained from this MCP were it operated in the saturated mode at 3 kV.

5.8 Multiplier and PMT gain measurement

Methods for the measurement of the two aspects of gain introduced in 5.1 are presented. Recall that multiplier gain, g , is defined with reference to a random source of single photons: it is the ratio of the anode charge to that of a photoelectron, when viewing a source of single photons. Chevrons are used to emphasize its

statistical nature and that we are dealing with mean gain. Multiplier gain is dimensionless since both parameters are expressed in coulombs. There is a second, dimensionless representation, G , known as ‘photomultiplier gain’ and defined as anode current divided by photocurrent. The expectation might be that the two formulations should lead to the same gain figure. However, gain based on current measurement is always lower than that given by the charge ratio because a fraction of the photocurrent fails to propagate, whereas charge pulses recorded at the anode originate only from those photoelectrons that have successfully propagated. Here, F is known as the ‘collection efficiency’, and the three parameters are related thus: $F = G/g$; F is a fundamental attribute of PMTs, and it is investigated in its own right in 10.3. It is important to illuminate any PMT under investigation in the same way as intended in the application, noting that gain measurements are sensitive to wavelength, angle of incidence, and area of illumination.

5.8.1 Measurement of G : First method

The quotient of anode current and cathode current is the basis of PMT gain, G . In theory, it is determined by simultaneously measuring cathode current with a picoammeter, and anode current with a micro-ammeter. However, measurement of PMT gain presents a challenge where high gain is concerned. The range in anode current must be restricted to avoid non-linear operation. There are two such sources: one originates in the PMT, while the other arises from the action of the voltage divider. The problem is immediately obvious if we consider an attempt to measure G of the order of $\sim 10^7$ with a PMT operating in a resistive divider. To avoid fatigue and non-linear operation, the anode current should not exceed $10 \mu\text{A}$, implying a photocurrent of the order of 1 pA . Measuring such a small current with any certainty is unreliable, particularly in the presence of signal induced leakage currents discussed by Young (1971). He quotes Tothill of EMI: ‘If attempts are made to measure cathode currents in the region of 10^{-11} A , when the tube is operating with a substantial anode current, a spuriously high photocurrent is usually measured’.

The method, however, is acceptable for the measurement of low gain, less than 10^4 , with a photocurrent of nanoamps (although photocurrent non-linearity will need checking at the relatively high light levels involved). Gain determination beyond 10^4 can be achieved by increasing the voltage in stages, starting at the initial condition (G_0, V_0) , where $G_0 = 10^4$. The first step is to reduce the anode current by a factor of 10 by adjusting the light level only; the voltage is then increased until the original anode current is attained, arriving at the condition (G_1, V_1) , where G_1 is now 10^5 , and V_1 is the corresponding voltage setting. This procedure is repeated, using the minimum number of steps, until the desired gain is achieved. The expectation is that the systematic errors would be minimal, but it is not clear how to ascertain their contribution. Note that the method also provides the means to generate a gain–voltage curve, although finer steps may be beneficial.

5.8.2 Measurement of G: Second method

This method bears similarities to the procedure in 5.8.1 in attempting to overcome the problems associated with low cathode currents. An analogous scaling problem arises in the use of neutral density (ND) filters in the methods adopted by Coates (1973), Lakes and Poultney (1971), and Young and Schild (1971). Central to these investigations into measuring I_a is the initial choice of a relatively high cathode current, with the PMT operating either as a diode or limited to three dynode stages only. Optical filters, calibrated in situ for transmission, are introduced and the PMT is ultimately restored to its fully active state. In effect, gain scaling uncertainties, adopted in the first method, have simply been replaced by light attenuation ones, involving factors ranging from 10^4 to 10^9 . The authors of these investigations list the following hazards with the use of filters: infrared transmission, spatial non-uniformity, light leaks, interaction between filters, and pinholes. Anyone contemplating the use of optical filters in this way is advised to consult these publications first.

5.8.3 Measurement of G: Third method

This method for determining G is based on photocathode and anode sensitivity data supplied by manufacturers. The appropriate parameters are photocathode radiant sensitivity and anode radiant sensitivity, when considering photons of a single wavelength, or light constrained within a narrow wavelength band. The spectral responsivity $\sigma_k(\lambda)$ for a particular photocathode is directly related to the QE $\eta(\lambda)$, at the given wavelength. The units of $\sigma_k(\lambda)$ are millamps per watt, with λ is expressed in nanometres (see 2.3):

$$\sigma_k(\lambda) = \lambda\eta(\lambda)/1.24 \text{ mA/W}. \quad (5.14)$$

As an example, a typical bialkali photocathode has $\eta(\lambda) = 0.25$ at 400 nm, which is equivalent to a radiant sensitivity of 81 mA/W. The anode output, assuming a multiplier gain of G , is

$$\sigma_a(\lambda) = G \times \sigma_k(\lambda) = 81 \times 10^3 \text{ A/W} \quad (5.15)$$

for $G = 10^6$, for example. The relationship in (5.15) emphasizes the high sensitivity of PMTs and explains why laser applications generally use devices of less than 1 nW in power; G is the ratio of $\sigma_a(\lambda)$ and $\sigma_k(\lambda)$ and it may be provided by a manufacturer, for a stated voltage. In effect, the manufacturer has already measured the PMT gain. Cathode luminous sensitivity is denoted by S_k (expressed in microamps per lumen), and the anode sensitivity (expressed in amps per lumen) is defined in terms of the lumen and the PMT gain, G .

$$S_a = S_k \times G. \quad (5.16)$$

Manufacturers quote both cathode and anode luminous sensitivity figures for the majority of PMTs. VUV and solar blind types are the exceptions because they do not respond to white light and are better specified in terms of wavelength response.

A typical test ticket data set for a 30 mm bialkali PMT is

Cathode luminous sensitivity, S_k :	$112 \mu\text{A/lm}$
Voltage for anode sensitivity of 200 A/lm:	910 volts
Voltage for anode sensitivity of 2000 A/lm:	1220 volts

We can determine G from (5.16) at the two quoted voltages: G is dimensionless and equal to $200/(112 \times 10^{-6}) = 1.79 \times 10^6$ at 910 V, and to 1.79×10^7 at 1220 V. These data allow us to construct a gain–voltage curve based on the semi-empirical form $G = cV^\beta$, with:

$$\beta = \frac{\log\left(\frac{1.79 \times 10^7}{1.79 \times 10^6}\right)}{\log\left(\frac{1220}{910}\right)} = 7.85$$

$$c = \frac{1.79 \times 10^7}{1220^{7.85}} = 1.025 \times 10^{-17}$$

5

Do not be alarmed because c is so small: it is counter-balanced by V^β . A gain–voltage table is given in Table 5.1 and a useful check is to confirm that the curve passes through the calibration points, shown in bold.

Table 5.1 Gain–voltage table derived from two spot readings, provided by the manufacturer's test ticket, using the relationship $G = cV^\beta$.

Voltage (V)	G	Amps per lumen
700	2.26×10^5	25
750	3.89×10^5	43
800	6.46×10^5	72
850	1.04×10^6	116
900	1.63×10^6	182
910	1.79×10^6	200
950	2.50×10^6	279
1000	3.74×10^6	418
1050	5.49×10^6	614
1100	7.92×10^6	885
1150	1.12×10^7	1256
1200	1.57×10^7	1756
1220	1.79×10^7	2000
1250	2.17×10^7	2421

The question arises as to the validity of this tabular information and its accuracy. This is difficult to answer, primarily because of the particular way in which manufacturers determine cathode and anode sensitivities, and the reliance on filters of 'known' attenuation. Problems encountered in 5.8.1 and 5.8.2 apply equally to manufacturers. The data in Table 5.1 refer to a particular voltage divider configuration, usually stated by the manufacturer, and the table therefore only applies if the same voltage distribution is subsequently chosen by the user. A further complication lies in the choice of illumination area: ET Enterprises, for example, state in their literature that 80% of the photocathode area is illuminated in their test, a condition not necessarily replicated by a user. An accurate knowledge of the PMT gain G is arguably important only where absolute light measurements are attempted. For applications involving multiple tube usage, with deliveries spread over time, the important consideration is consistency in manufacturers' data. Such users may rely on test ticket information for grading, selection, and replacement purposes, where clearly consistency is a key factor. There remains the difficult and debatable question concerning accuracy: the author's belief, not necessarily shared by manufacturers, is that PMT gain determination by the methods of this section should be regarded as order of magnitude only.

5.8.4 Measurement of $\langle g \rangle$: First method

This method is based on an application of the shot noise formula to the photocurrent, with subsequent transformation to the anode. Signal is taken as $I_k F$, allowing for the loss in photocurrent attributed to finite collection efficiency, F . Shot noise relates to $I_k F$ but not to I_k alone, and the signal-to-noise ratio is

$$S_k / N_k = I_k F / (2e I_k F B)^{1/2}. \quad (5.17)$$

The signal-to-noise ratio at the anode is

$$\begin{aligned} S_a / N_a &= \frac{I_k F \langle g \rangle}{(2e I_k F B)^{1/2} \langle g \rangle} \\ &= \frac{I_k F}{(2e I_k F B)^{1/2}} \end{aligned} \quad (5.18)$$

Note that $\langle g \rangle$ drops out of equation (5.18) because the expression is ratiometric, and similarly for S_k / N_k . By definition $I_k F = I_a / \langle g \rangle$, and

$$(S_a / N_a)^2 = \frac{I_a}{2e B \langle g \rangle}. \quad (5.19)$$

A plot of $(S_a / N_a)^2$ versus I_a gives a straight line of slope $s = 1/(2e B \langle g \rangle)$ and it follows that

$$\langle g \rangle = \frac{1}{2e B s}. \quad (5.20)$$

The value of $(S_a/N_a)^2$ can be determined experimentally using a DMM to measure V_{DC}/V_{rms} , which is identical to S_a/N_a provided a true rms meter is used. The required bandwidth is derived from the combination of B_{DMM} , given by the manufacturers of the DMM as 2000 Hz, and $B_{RC} = 1/(4RC)$, where R is the parallel combination of the load resistor and the DMM resistance, and C is the total capacitance in the circuit. Note the appropriate bandwidth to use is the noise bandwidth $1/4RC$ and not $1/2\pi RC$, as is often wrongly assumed. Bandwidths combine in quadrature to give an overall bandwidth of

$$B = B_{DMM} \times B_{RC} / (B_{DMM}^2 + B_{RC}^2)^{1/2}.$$

In the example given, B is determined as 963 Hz, and the slope extracted from Fig. 5.40 is 1.29×10^8 , leading to a $\langle g \rangle$ of 2.5×10^7 . This figure is an overestimate of $\langle g \rangle$ because multiplier noise has not been included. A noise factor from (4.77) is readily included in (5.19), as follows:

$$(S_a/N_a)^2 = \frac{I_a}{(2eB\langle g \rangle) \left(1 + \frac{\text{var}(g)}{\langle g \rangle^2} \right)}. \quad (5.21)$$

The noise factor for this particular PMT is ~ 1.25 which, when included, reduces $\langle g \rangle$ to 2.0×10^7 ; this is in reasonable agreement with the figure obtained using an MCA, as described in 5.8.5.

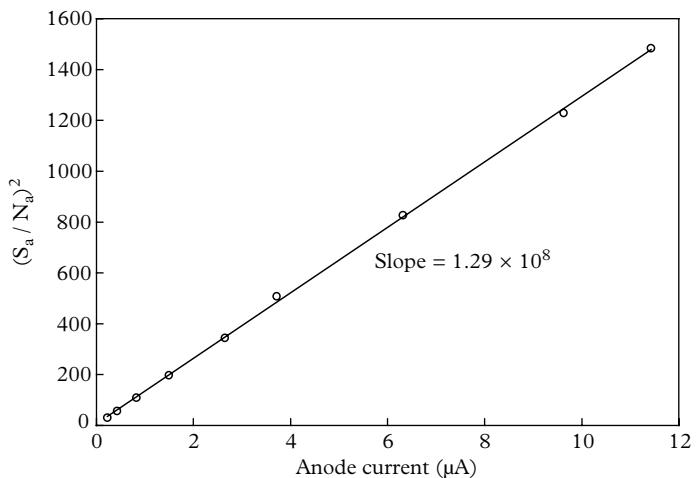


Fig. 5.40. Gain may be determined based on the shot noise formula. The effective resistance is the load resistance in parallel with the impedance of the DMM, that is, $1\text{M}\Omega \parallel 10\text{M}\Omega$. The total capacitance is 250 pF, comprising the capacitance of the cable and the PMT assembly and readily measured with a capacitance meter.

5.8.5 Measurement of $\langle g \rangle$: Second method

Multiplier gain, $\langle g \rangle$, refers to the relationship between the charge developed at the anode, q_a , generated by a charge, q_k , at the photocathode. If q_k is taken as the electronic charge e , the multiplier gain is given by

$$\langle g \rangle = \langle q_a \rangle / e. \quad (5.22)$$

A single-photon light source is easily devised in the manner described in 4.1.1 and, since $\langle q_a \rangle$ is statistical in nature, we are dealing with an output pulse height distribution in charge, depicted in Fig. 5.41. An MCA incorporating a charge sensitive preamplifier is an ideal instrument for quantifying q_a . The calibration is 3.2×10^{-14} Coulombs/channel for the particular arrangement illustrated; calibration is based on applying a known voltage step to a small capacitor connected to the preamplifier input. Gain is directly related to the mean of the pulse height distribution (Channel 94.4), $\langle q_a \rangle$ expressed in Coulombs divided by the electronic charge:

$$\langle g \rangle = 3.2 \times 10^{-14} \times \frac{94.4}{1.6 \times 10^{-19}} = 1.89 \times 10^7.$$

The accuracy of this method depends upon the reliability of the adopted charge calibration procedure, and also on the uncertainty in determining the mean of the pulse height distribution; particularly the way in which it is affected by the

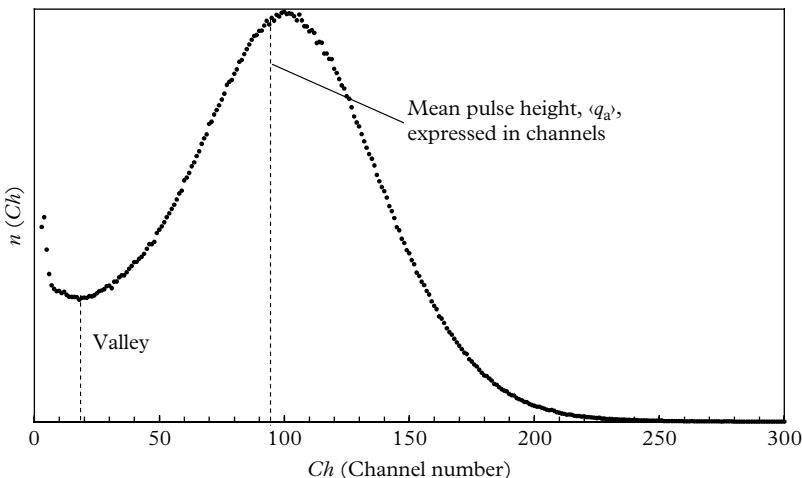


Fig. 5.41. The output charge distribution, expressed in channel numbers, from a 9214B PMT excited with single photons. This distribution is the SER. Channel numbers can be expressed in picocoulombs by using the calibration factor 3.2×10^{-14} coulombs/channel. Note the peak (the most probable gain) occurs at Channel 100, and the mean at Channel 94.4, in this example. Channel 3 is the lowest one, containing only signal counts.

inclusion, or otherwise, of the sharply rising low channel counts. This difficulty also applies to the next method for gain determination. The important question concerning the accuracy in determining $\langle g \rangle$ is not so easily answered. Calculation should include all channels from 1 to 1024 but for this instrument readings start from the third channel because of offset in the analyser. One way in which to explore the sensitivity to the presence of small pulse heights is to examine how $\langle g \rangle$ is affected when small pulse height counts in Channel 1 and Channel 2 are omitted from the calculation; this procedure increases the mean by only $\sim 1\%$. There is also a 1% uncertainty in the calibration of the charge sensitive amplifier (dictated by the tolerance of the charge injection capacitor).

5.8.6 Measurement of $\langle g \rangle$: Third method

In this method, $\langle g \rangle$ is determined by integration of the SER, such as the one shown in Fig. 5.41, giving a total count rate, N . If the corresponding anode current, I_a , is measured simultaneously, then

$$I_a = N e \langle g \rangle, \quad (5.23)$$

where N is the sum of the channel counts, expressed as a rate per second. The challenge of including all pulse heights in N is the same as for the previous method, and is illustrated in Fig. 5.42; N may be deduced by extrapolation based on Channels 3 and 4, as illustrated, and gain, $\langle g \rangle$, then follows from (5.23). As in the previous method, only those photoelectrons that propagate to the anode

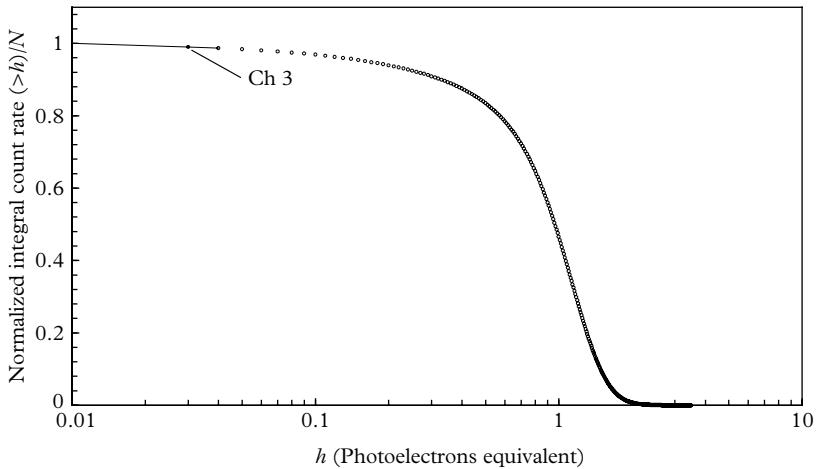


Fig. 5.42. The integral pulse height distribution derived from Fig. 5.41, after normalizing to unity. The extrapolation shown gives N , the total count rate. The inclusion of a linear extrapolation from $h = 0.03$ pe (Channel 3) to $h \rightarrow 0$ increases N by only 1.4%. This is indicative of the accuracy of the method.

contribute to both I_a and N . Those that are lost, in accordance with the finite collection efficiency, make no contribution and hence $\langle g \rangle$ is independent of F , the collection efficiency—as it must be. Dead time does not affect the measurement of I_a but N requires correction. The count rate should in any case be sufficiently low to ensure the magnitude of the correction is only a few per cent. MCAs are designed to extend the run duration to compensate for ‘busy’ time, but the dead time correction is indicative only. Dark count and dark current may be subtracted to yield $\langle g \rangle$ for signal alone, but this is unnecessary where the dark count and signal SERs are similar in shape.

5.8.7 Measurement of $\langle g \rangle$: Fourth method

Detailed pulse height information is not actually necessary for determining gain by this method: the requirements are the total count, N , and hence the count rate derived by using a threshold discriminator is sufficient. The gain, $\langle g \rangle$, is determined from (5.43) using anode current and corresponding counts measurements. Anode current may be determined from a voltage measurement by switching in a high impedance load—typically 100 k Ω . Amplifier-discriminators, however, are characterized by a low counting threshold of a few millivolts and an input impedance of 50 Ω . Consequently, the mismatch in impedances rules out simultaneous measurement of I_a and N ; readings are made by successively switching in the high impedance load for measurement of I_a . The method is suitable for both HV polarities, but positive HV demands connection of the load resistor to the HV supply; I_a is determined from a voltage reading, using the expedient of floating a multimeter, as already discussed. The choice of negative HV provides the option of measuring anode current directly. This method of gain determination highlights the problem already encountered in 5.8.5: what count rate corresponds to the true multiplier gain? Arguably, based on Fig. 5.43, all that can be safely deduced is that N lies between 30,000 and 40,000 counts/s, leading to an uncertainty of about 30% in determining $\langle g \rangle$ by this method. The source of the problem is that the ‘plateau’ characteristic, despite the established terminology, slopes with increasing voltage. This leads to the conclusion that the integral rate cannot be accurately determined from a plateau characteristic (see A.11). The gain, at a stated operating voltage within the range of the PMT, is calculated using the measured anode current and N at that operating voltage. The merit of this method lies in its ease of measurement in providing an order of magnitude determination of gain, but it is unsuitable where absolute gain is required.

5.9 Quick determination of $\langle g \rangle$

Certain measurement techniques involve the use of an MCA, offering accuracy of about 5%, while other techniques only need a digital oscilloscope and a laboratory

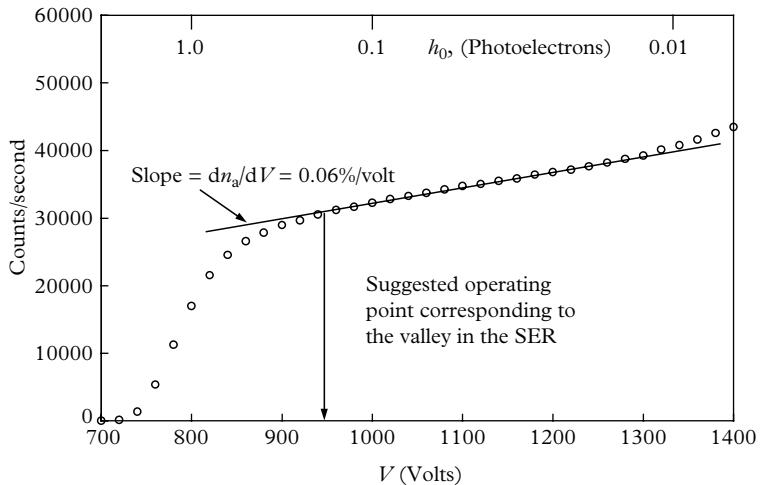


Fig. 5.43. Plateau characteristics measured with an ET Enterprises AD2 amplifier-discriminator operating with 5 μ s dead time. The slope of the characteristic depends on the quality of the SER, the change in collection efficiency as V advances, backscatter contributions, and the presence of afterpulsing.

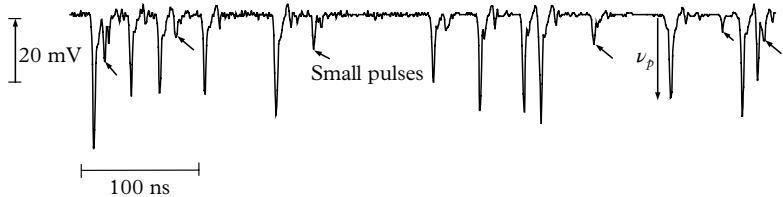


Fig. 5.44. An oscilloscope trace of single-photon events, showing a typical height of 25 mV, with $t_{\text{fwhm}} \approx 2.5$ ns, measured by expanding the time base. Small pulses are also clearly evident. There is uncertainty in distinguishing small pulses from ringing on the larger ones but this could be mitigated by improved matching.

multimeter. The virtue of adopting a simple approach is one of providing immediate results, albeit at a lower level of certainty. The primary purpose for introducing these measurement techniques is to provide a rough determination of $\langle g \rangle$, perhaps as a check on the methods already described. Quick methods are attractive because of their simplicity, and the time necessary to produce a reasonable estimate of $\langle g \rangle$. In addition, the instrumentation required is minimal: a digital oscilloscope, a multimeter, and a source of single photons.

5.9.1 Quick determination of $\langle g \rangle$: First method

Multiplier gain is deduced from the following relationships, in which q_a is the charge at the anode initiated by a single photoelectron. This is a statistical quantity related to the gain, with an expectation, $\langle q_a \rangle$, obtained by taking sufficient single-shots. The effective duration of the pulse, τ , is taken equal to its fwhm height, t_{fwhm} , resulting in a pulse of peak voltage $\langle v_p \rangle = \langle q_a \rangle R / \tau$. We have

$$\begin{aligned}\langle g \rangle &= \langle q_a \rangle / e \\ \langle q_a \rangle &= \langle i_p \times \tau \rangle = \langle v_p \rangle \tau / R \\ \langle g \rangle &= \langle v_p \rangle \times \frac{t_{\text{fwhm}}}{R e},\end{aligned}\quad (5.24)$$

where $i_p \tau$ is the charge produced at the anode by the detection of a photoelectron. The peak voltage v_p and the fwhm height t_{fwhm} are measured with a fast, 50Ω input-impedance oscilloscope. The requirement in equation (5.24) is that R be sufficiently small to reduce the effect of stray capacitance in the calculation of $\langle g \rangle$ ($R = 50 \Omega$ certainly meets this requirement). A typical single-shot oscilloscope trace in Fig. 5.44 is a sample of the pulse height distribution, from which a gain of $\sim 10^7$ follows.

An alternative treatment of the data in Fig. 5.44 involves counting pulses and noting the corresponding anode current. There are 17 such pulses, representing a rate of 4.5 MHz, and a multimeter indicates an anode current, I_a , of $7.1 \mu\text{A}$. Multiplier gain predicted by equation (5.24) is thus $\sim 10^7$. The precision depends on the number of shots taken but the accuracy is sensitive to the inclusion, or otherwise, of small pulses—clearly a matter of judgement.

5.9.2 Quick determination of $\langle g \rangle$: Second method

This method is based on the pulse height attained by a single-photoelectron pulse captured by a capacitor, where

$$\langle g \rangle = C \langle v_p \rangle / e. \quad (5.25)$$

In this example, a capacitance of 90 pF includes a short length of coax cable, stray capacitance associated with the anode, and the input capacitance of the oscilloscope (set to $1 \text{ M}\Omega$ input impedance). A $56 \text{ k}\Omega$ resistor is added in parallel to provide a time constant of $\sim 5 \mu\text{s}$. This is considerably longer than τ (2.5 ns) and hence total integration of every input charge is ensured. Individual pulse heights, v_p , are read off the trace in Fig. 5.45 for averaging, and the mean is substituted into equation (5.25). As previously, $\langle g \rangle$ turns out to be of the order of 10^7 , but the accuracy is considerably improved over the first method, as uncertainty lies only in the acceptance threshold for small pulses. A measurement of anode current and frequency also provides a measurement of $\langle g \rangle$ with this arrangement.

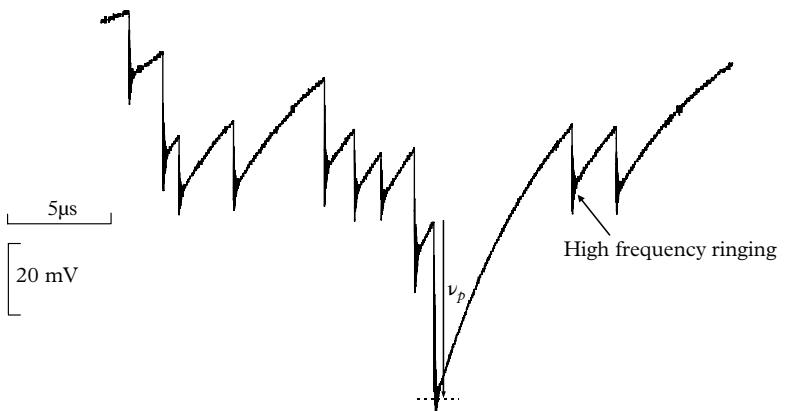


Fig. 5.45. Oscilloscope traces for the determination of $\langle g \rangle$ based on the voltage developed across a known capacitance. Ringing at the base of each step could be eliminated by using a shorter coaxial cable, or judicious bandwidth limiting.

5.10 Gain-voltage characteristics

The best method for making a multiplier gain measurement is the one based on the SER. In practice, this means operating a device at the relatively high gain of $\sim 10^6$. At this gain, the mean output charge generated at the anode is 0.16 pC, which is comfortably within the sensitivity range of most MCAs. The charge sensitivity of an MCA can be established in terms of picocoulombs per channel, by applying a known voltage step at the preamplifier input (see 5.8.5). The most probable gain follows immediately for those distributions that exhibit a peak, and the mean gain can be calculated by straightforward manipulation of the MCA data, whether there is a peaked response or not. This method provides a measure of gain at the set HV. A gain–voltage characteristic may be determined, limited by the dynamic range of the MCA, by noting the SER peak position as the HV is varied. Alternatively, where negative HV is used, the measured variation of the anode current with voltage, V , provides the basis for generating a gain–voltage curve, the peak or the mean of the SER serving as the normalization point for such readings. Scaling is necessary to ensure that the PMT maintains linearly. The light level may be increased at low voltage settings, again using the PMT itself to quantify the numerical shift.

It will be found that relative gain change with HV is sufficient for most purposes. This is certainly the case in transforming distributions from one variable to another (see A.1), where only a knowledge of the log slope, β , is required. PMT manufacturers usually supply ticket information for two gain settings from which a gain–voltage characteristic is readily constructed, in the manner of 5.8.3.

A sensitivity–voltage curve for a PMT, as opposed to that for a multiplier, is sometimes measured in terms of amps per lumen or milliamps per watt but it is likely that only someone in a standards laboratory, or employed by a PMT manufacturer, would have the necessary equipment, or inclination, to do so.

5.11 Conclusions

Sections 5.8 and 5.9 reveal the enigma of PMTs: gain is finite and yet elusive to specify with total confidence. PMT gain would appear amenable to reliable measurement. However, the significant difference in magnitude between the two currents I_a and I_k poses experimental uncertainties in the inevitable resort to light attenuation. A further problem concerns photocurrent because a certain percentage fails to propagate to the anode. In theory a correction can be made in terms of collection efficiency, F , but this has its own computational uncertainties.

The accuracy of the methods for determining multiplier gain depends primarily on the degree to which small pulses are included in the calculation. Although these pulses make only a small contribution to I_a , they are counted with the same weight as any other pulse. There is often a significant contribution from small pulses in a background pulse height distribution, which is explained by electron emission from the dynodes. However, there are also small signal pulses present when viewing any light source, but it is known that these contributions are signal or signal induced because they disappear on removal of the light signal. In all methods based on count rate and anode current, there is a tendency to overestimate $\langle g \rangle$ since afterpulses contribute to I_a but not necessarily to counts, N —this contribution depends on dead time and is particularly relevant in the use of an MCA with a dead time of a microsecond or more.

The mean gain of a PMT is generally lower than the peak gain, and the shape of the SER determines its location. The most probable gain has no obvious theoretical significance; however, it has practical application in aligning PMTs that exhibit a peaked SER. The peak and the mean channel numbers are typically within $\sim 10\%$ of one another and it is usually sufficient and more convenient to use peak gain, thereby avoiding detailed computation. Standard PMT alignment practice in high-energy and astrophysics experiments is to set all PMTs to the same peak gain, which, further to the argument just made, ignores the spread in photocathode sensitivity; this may be included in a look-up table in software. If all PMTs are of the same type, it is reasonable to assume they have a common F value, making this correction straightforward and yet uncertain.

Contributions as small as one-hundredth of a photoelectron equivalent are present in the SER of PMTs. These are attributed to a diminution in secondary-electron yield, and hence overall gain, caused by inelastic scattering. The shape of an SER also depends on whether the single-photon source used for measurement

is random or gated. Consequently, the gain is different in these two cases because an SER measured under gated conditions has fewer small pulses, which suggests a contribution from signal induced counts.

Referring to Fig. 5.43, the slope of the plateau characteristic leads to uncertainty in determining N . This applies primarily to the method described in 5.8.7, where counts are provided by an amplifier discriminator. It is clear from theoretical distributions in Fig. A.7 (Appendix A) that even at the low thresholds, $N(>h_0, V)$ only attains its final value asymptotically, when $V \rightarrow \infty$. In addition, the slope is enhanced by the presence of afterpulses, adding further uncertainty. In contrast, the configuration of Fig. A.6 does not suffer in this regard, as is evident by the convergence of the family of integral curves to a common ordinate. It is customary in the scientific and in manufacturers' literature to choose a log scale for the count rate (the ordinate), a ruse to justify the concept of a counting plateau, when in reality the count rate is better described as a region of minimal slope. Note that the characteristic for a SiPM approaches a true plateau as there is little backscatter in view of low Z_{eff} . Use of a discriminator in the method of 5.8.7 can be successfully employed by measuring N as a function of threshold setting at fixed V . The threshold can be changed either by altering the amplifier gain, changing the discriminator threshold, or by including a $50\ \Omega$ stepped attenuator between the anode and the discriminator (see Fig. A.1).

It is sensible to ask, under what circumstances is knowledge of gain essential? Clearly, this is important at the design and prototype stage for manufacturers. Users involved in making absolute measurements of light will need to know QE, collection efficiency, and gain. All these characteristics carry uncertainty, making sensitivity predictions possibly no better than 20%. A higher level of accuracy can only be achieved by using a calibrated light source, or a calibrated detector, from a national standards laboratory—the options are discussed in 2.3.

REFERENCES

- Abbon, P. *et al.* (2008). The characterisation of the multianode tubes for RICH-1 upgrade project at compass. *Nucl. Instr. and Meth. in Phys. Res. A*, **595**, 177–9.
- Agoritsas, V., Kuroda, K. and Nemoz, Ch. (1989). A new type of position-sensitive electron multiplier. *Nucl. Instr. and Meth. in Phys. Res. A*, **277**, 237–41.
- Allen, J. S. (1950). Recent applications of electron multiplier tubes. *Proc. IRE*, **38**, 346–58.
- Baldwin, G. C. and Friedman, S. I. (1965). Statistics of single-electron multiplication. *Nucl. Instr. and Meth.*, **36**, 16–18.
- Chirikov-Zorin, I. *et al.* (2001). Method for precise analysis of the metal package photomultiplier single photoelectron spectra. *Nucl. Instr. and Meth. in Phys. Res. A*, **456**, 310–24.

- Coates, P. B. (1970a). The edge effect in electron multiplier statistics. *J. Phys. D: Appl. Phys.*, **3**, 1290–6.
- Coates, P. B. (1970b). The gain of gallium phosphide dynodes. *J. Phys. D: Appl. Phys.*, **3**, L25–L27.
- Coates, P. B. (1971). Noise sources in the C 31000 D photomultiplier. *J. Phys. E: Sci. Instr.*, **4**, 201–7.
- Coates, P. B. (1973). Photomultiplier collection efficiencies and nonpoissonian pulse height distributions. *J. Phys. D: Appl. Phys.*, **6**, 153–63.
- Darlington, E. H. (1975). Backscattering of 10–100 keV electrons from thick targets. *J. Phys. D: Appl. Phys.*, **8**, 85–93.
- Dekker, A. J. (1958). Secondary electron emission. *Solid State Phys.*, **6**, 251–311.
- Delaney, C. G. F. and Walton, P. W. (1964). Single electron spectra in photomultipliers. *Nucl. Inst. and Meth.*, **25**, 353–6.
- Dietz, L. A., Hanrahan, L. R., and Hance, A. B. (1967). Single electron response of a porous KCl transmission dynode and application of Polya statistics to particle counting in an electron multiplier. *Rev. Sci. Instr.*, **38**, 176–82.
- Evans, R. D. (1955). *The Atomic Nucleus*. McGraw-Hill, New York.
- Gutierrez, W. A., Pommerenig, H. D., and Holt, S. L. (1972). Secondary electron emission from GaAs. *Appl. Phys. Lett.*, **21**, 249–250.
- Klettee, B. D., Krym, N. D., and Wolber, W. G. (1970). Long-term stability characteristics of commonly used channel electron multipliers. *IEEE Trans. NS.*, **17**, 72–80.
- Kuroda, K., Sillou, D., and Takeutchi, F. (1981). New type of position sensitive photomultiplier. *Rev. Sci. Instr.*, **52**, 337–46.
- Kyushima, H. et al. (1994). Photomultiplier tube of new dynode configuration. *IEEE Trans. NS*, **41**, 725–9.
- Lakes, R. S. and Poultney, S. K. (1971). Direct measurement of the quantum counting efficiency of RCA C31000E/F photomultipliers at 6328 Å. *Appl. Opt.*, **10**, 797–800.
- Martinelli, R. U., Schultz, M. L., and Gossenberger, H. F. (1972). Reflection and transmission secondary emission from GaAs. *J. Appl. Phys.*, **43**, 4803–4.
- Nilsson, O. et al. (1970). Development of parallel plate channel multipliers for use in electron spectroscopy. *Nucl. Instr. and Meth.*, **84**, 301–6.
- Oliver, C. J. and Pike, E. R. (1970). The edge effect in electron multiplier statistics. *J. Phys. D: Appl. Phys.*, **3**, L73–L74.
- Pietri, G. (1975). Towards picosecond resolution. Contribution of microchannel electron multipliers to vacuum tube design. *IEEE Trans. NS*, **22**, 2084–92.
- Prescott, J. R. (1966). A statistical model for photomultiplier single-electron statistics. *Nucl. Instr. and Meth.*, **39**, 173–9.
- Ray, J. A. and Barnett, C. F. (1970). Characteristics of a funnel type electron channel multiplier with a grid entrance window. *IEEE Trans. NS*, **17**, 44–8.
- Reynolds, G. T. (1966). The distribution of single electron pulse sizes from multidynode electron multipliers, and single electron detection. *Adv. Electron. El. Phys.*, **22A**, 71–85.

256 Secondary emission and gain

- Schmidt, K. C. and Hendee, C. F. (1966). Continuous channel electron multiplier operated in the pulse saturation mode. *IEEE Trans. NS*, **13**, 100–11.
- Simon, R. E. and Williams, B. F. (1968). Secondary-electron emission. *IEEE Trans. NS*, **15**, 166–70.
- Slark, N. A. and Woolgar, A. J. (1962). Transmission secondary emission image intensifiers. *IEEE Trans. NS*, **9**, 115–17.
- Sommer, A. (1972). Bialkali (K_2CsSb) photocathodes as a high gain secondary electron emitter. *J. Appl. Phys.*, **43**, 2479–80.
- Stuchiinsky, G. B. (1970). Inelastic electron reflection from emitters of Cs_3Sb and $(Cs)Na_2KSb$. *Radio Eng. and Electr. Phys.*, **15**, 360–1.
- Thomas, S. and Pattinson, E. B. (1970). Range of electrons and contribution of back-scattered electrons in secondary production in aluminium. *J. Phys. D: Appl. Phys.*, **3**, 349–57.
- Wiza, J. L. (1979). Microchannel plate detectors. *Nucl. Instr. and Meth.*, **162**, 587–601.
- Young, A. T. (1971) Use of photomultiplier tubes for photon counting. *Appl. Opt.*, **10**, 1681–3.
- Young, A. T. and Schild, R. E. (1971). Photoelectron collection efficiency in photomultipliers. *Appl. Opt.*, **10**, 1668–72.
- Young, J. R. (1956). Penetration of electrons in aluminium oxide films. *Phys. Rev.*, **103**, 292–3.
- Zworykin, V. J. and Rajchman, J. A. (1939). The electrostatic electron multiplier. *Proc. IRE*, **27**, 558–66.

6

PMT background

6.1 Introduction

The term background refers to the signal at the output of a PMT when operated in complete darkness. Obtaining minimal and stable background is particularly important in ultra-low-light-level applications—typically where rare events are involved. Indeed, major astrophysics experiments of recent times would not have been possible without the development of PMTs with particularly low background. The presence of PMT background counts manifests itself primarily under high-gain conditions—typically greater than 10^6 . The reason for this is that the background spectrum consists mainly of single-photoelectron events; there is a characteristic shape to the background spectrum, depending on the photocathode type, the diameter of the PMT, the window material, and its thickness. Background referred to as signal-induced is distinct from that covered in this chapter. The correlated form of background, discussed in Chapter 11, is generated by signal and is particularly important in applications involving timing of events.

It is well known and certainly publicized by all manufacturers, that a period of dark adaptation invariably results in background reduction, but it never disappears completely, no matter how long one waits. PMT background generally decreases with time; particularly under continuous operation and in the early days experienced users knew to switch on a week or more prior to final commissioning. But nowadays, manufacturers ‘hot-age’ products, for typically 12 h, to provide customers with PMTs that are substantially ‘run in’. It is assumed in this chapter that PMTs are fully dark-adapted.

An important point to bear in mind is that manufacturers are primarily concerned with meeting a specification ‘to get product out the door’, which, in the case of dark current, implies that a reading taken by a customer some months later, and after a period of continuous operation, will invariably be considerably lower than stated on the test ticket. Manufacturers only need to know that a given PMT meets the maximum dark current specification, but not by how much. The author’s experience is that a significant portion of PMTs taken out of stock for retest, say, 1 year after manufacture, show an improvement by a factor of 2 or more in their dark current figures. Manufacturers sometimes take advantage of this when supplying product to a special selection on dark current or dark count, especially if the number required is small. The improvement in dark current, in the way mentioned, does not apply to other test parameters, such as cathode sensitivity, where the measurements are generally repeatable to $\pm 5\%$ over the passage of time.

Consideration is given to the following categories of background:

- background from a PMT, which may emit light signals (crosstalk) and act as a source of background nuclear radiation

- background from the reverse situation, in which the environment is a source of natural radiation that affects PMT performance; for example, the natural gamma and cosmic ray backgrounds
- background from an internal source of signal generated by thermal emission from the photocathode
- signal-induced contributions, such as latepulses and afterpulses

It is apposite to regard sources of background simply as unwanted signal.

6.2 Dark counts and dark current

There are two ways to describe background: dark current and dark count, with the interpretation of ‘dark’ already defined. The former refers to a measurement with an electrometer, for example, connected in series with the anode. Dark current, with a magnitude typically within the nanoamps range, is primarily a function of gain. On the other hand, dark count rates are noticeably independent of gain and hence HV, provided that operation is confined within the manufacturer’s specification. The term ‘dark count’ refers to a measurement made with a discriminator–scaler combination, a digitizer, or an MCA. The result of such measurements is simply a number: the dark count rate for pulses with amplitudes that exceed a set threshold. Expanding on this statement, dark current scales approximately with gain, but dark counts scale inversely with threshold. This statement is a gross simplification, but it is nevertheless worth remembering.

The two ways for describing PMT background must reconcile where reference is to the same device. If a set of PMTs, all of the same gain, is first graded with respect to dark current, and then graded in terms of dark counts, the two data groups will not necessarily align. For this reason, PMT manufacturers prefer to supply selected product based on either dark current or on dark counts, as required by the application. It is shown in 6.3 how the ranking anomaly is resolved by introducing a third parameter, I_L , the leakage current. Manufacturers do not, however, include this parameter as part of the standard test ticket data.

A dark count pulse height distribution is most conveniently quantified with a charge-sensitive MCA, which measures the total charge in each pulse and bins it accordingly. The dark count rate is determined by integrating the differential distribution, typically from 0.1 to 100 photoelectrons equivalent—the chosen limits are subjective, particularly the lower one. A background pulse height distribution is shown in Fig. 6.1, together with one for single-photon excitation (signal) for the same PMT. For convenience, the SER signal rate is generally set relatively high compared with the background count, thus avoiding the need for background subtraction. The signal and background in Fig. 6.1 are normalized to the peak channel of the SER.

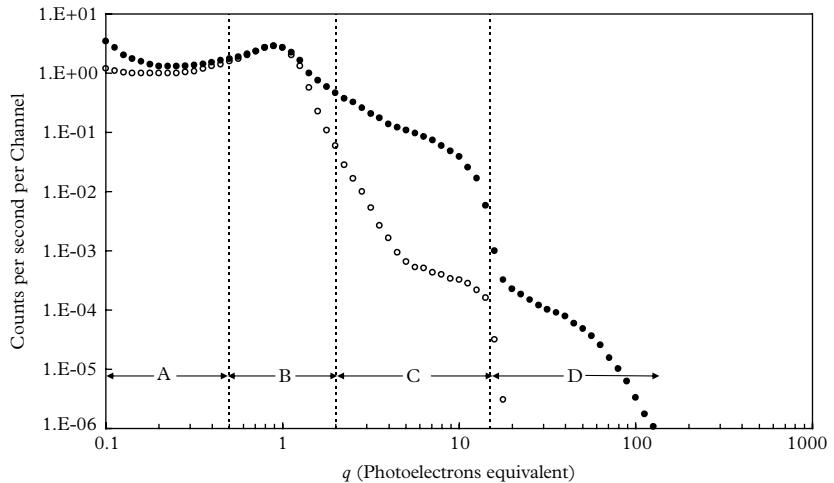


Fig. 6.1. Dark count pulse height (\bullet) distribution, measured for a 2" fast PMT exhibiting a well-defined single-electron peak; \circ represents the pulse height, under the same operating conditions, for single-photon excitation (signal).

6.2.1 Subdivision of a dark count distribution

The general features revealed in Fig. 6.1 are present in all PMTs. Hence, the abscissa may be divided into four pulse height bands: A, B, C, and D, with the boundaries defined in terms of photoelectrons, pe:

Region A: <0.5 pe

Region B: $0.5 < B < 2$ pe

Region C: $2.0 < C < 15$ pe

Region D: >15 pe

The immediate points to note regarding Fig. 6.1 are

- There are relatively more undersized events in Region A of the background than in the single-electron spectrum. This will be explained in terms of contributions to background from the dynodes.
- Region B is a measure of the single-photoelectron background.
- Region C has considerably more relative counts in the background compared with signal. Contributions are from afterpulses and radioactivity in the PMT window.
- Counts in Region D originate from natural gamma and cosmic ray sources.

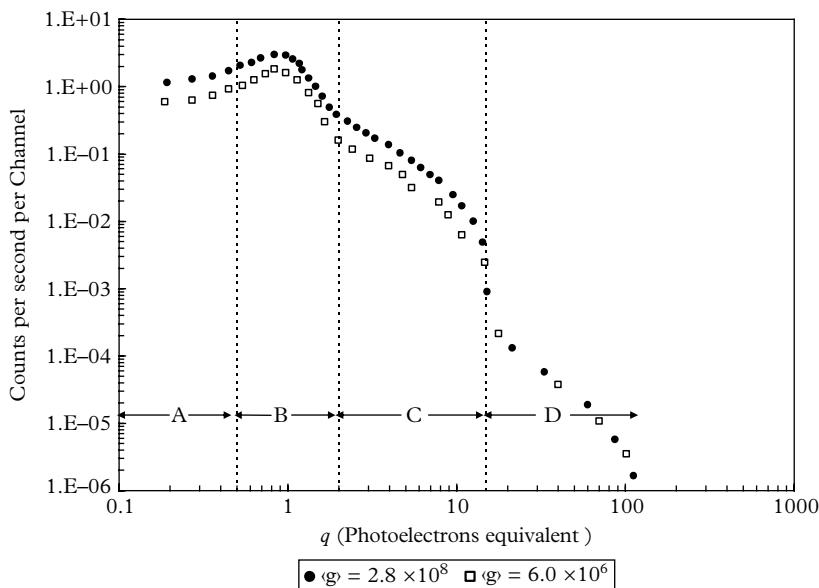


Fig. 6.2. Illustrating how the background pulse height distribution changes with gain. Note the distribution in Region D is unaffected.

The abscissa is expressed in photoelectrons equivalent by setting the peak of each distribution to unity. Alternatively, the abscissa may be expressed in pico-coulombs, if the PMT gain or the charge sensitivity of the MCA is known in terms of coulombs per channel. Figure 6.2 refers to a 50 mm diameter fast PMT and illustrates the small dependence on gain in all regions except D. Figure 6.3 shows the effect of temperature on a PMT with a quartz window. Total dark counts are ~ 10 per second at room temperature, reducing to ~ 1.5 at -25°C .

6

6.3 Reconciliation of dark current and dark counts

Dark current comprises two components: one pulsed and the other the result of leakage. Although care is taken in the design of a PMT, to insulate electrodes, the admission of alkali vapours at the processing stage has the effect of reducing the surface resistivity of the ceramic insulators. PMT pins, fused into a glass substrate, also take up alkali vapours. Applied biasing voltages generate leakage currents, of which those flowing to the anode, or to its pin, are of particular concern. In some cases, principally relevant to low dark count PMTs, the leakage current, I_L , may even exceed the pulsed component. The contribution from background pulses is

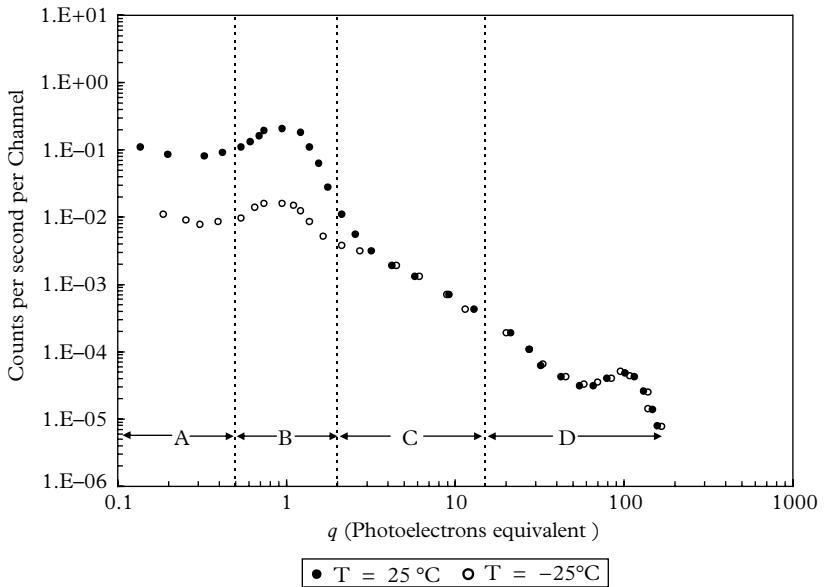


Fig. 6.3. An example of an exceptionally low-background quartz (9813Q) PMT. The integral counts in A and B decrease by a factor of 10, but C and D are temperature independent. Note the muon peak at 100 pe.

given by integration of a dark count spectrum, of the type shown in Fig. 6.1. The relationship between anode dark current, I_d count rate per channel, n_d and leakage current, I_L , is given by (6.1),

$$I_d = e g \int_0^{\infty} n_d(q) q dq + I_L, \quad (6.1)$$

and I_L may be estimated by connecting the cathode, first dynode, and second dynode together, while retaining the previous voltage distribution on the remaining dynodes. This effectively removes dark count contributions originating from the cathode and from the first dynode, which is easily checked by observing that there should now be very few output pulses. It is instructive to plot (6.1) taking the upper limit of the integral as q , to predict the relative contributions from Regions A through D. This has been done for the same PMT pertaining to Fig. 6.1, and is shown in Fig. 6.4. This curve is informative for it reveals that the small pulses in Region A contribute only 10% to the total, while those in C, although much fewer in number, generate about half the dark current.

The asymptotic value of i_q is $I_q = 2 \text{ nA}$ and this is the contribution to the dark current by the dark counts, at the stated gain. Results, covering a range of gains for the same PMT, are shown in Figure 6.5 as open circles. This represents the contribution to the dark current, I_d , by the first term on the right in (6.1).

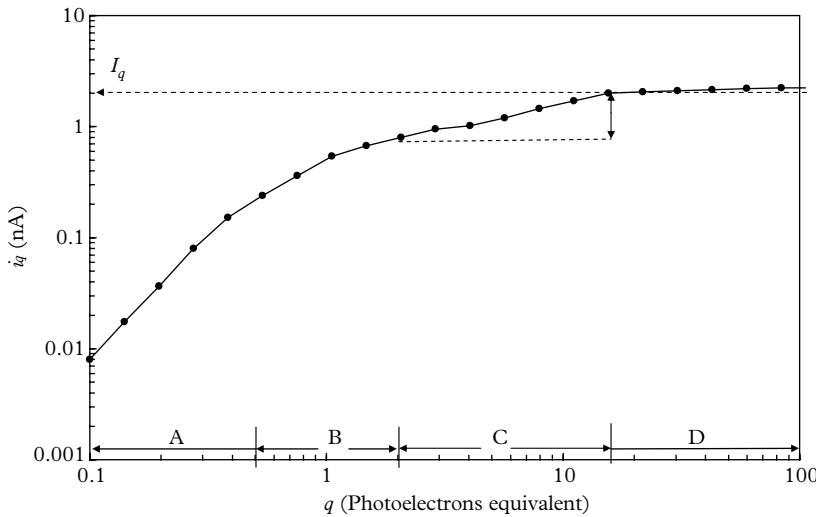


Fig. 6.4. The cumulative contribution of q to i_q from (6.1) at $\langle g \rangle = 5 \times 10^7$. The pair of arrows at 15 pe indicates the contribution made by the counts in C.

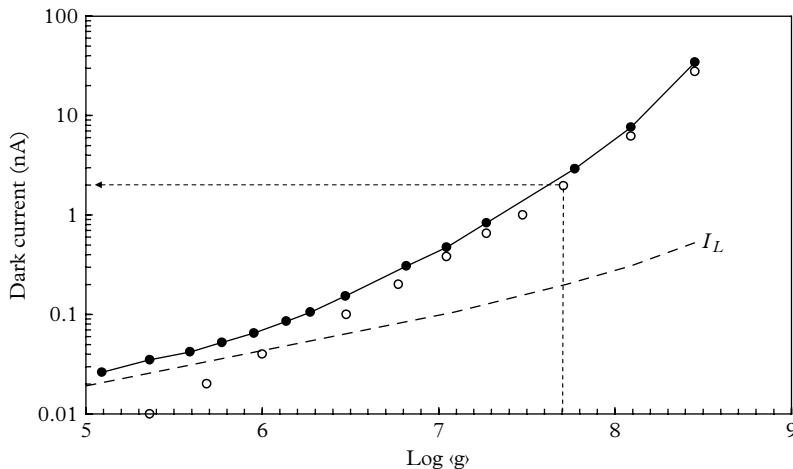


Fig. 6.5. Dark current as a function of gain is given by the solid dots. The predicted dark current, based on the sum of the dark count contribution (\circ) and the measured leakage current ($--$), is shown by the solid line.

Adding I_L to this contribution gives the predicted dark current, which may be compared with the actual measured dark current.

Some observations can be drawn from the measurements in Fig. 6.5, which apply to PMTs in general. The slope of the leakage current curve is shallower than the curve representing the dark count contribution because I_L is essentially a

function of HV, whereas the contribution from dark counts varies in proportion to the gain and hence as a high power of voltage. Leakage current dominates at gains below 10^6 , but its contribution is reduced at high gain. At ultra-high gain, say beyond 10^8 , the slope of the dark current curve increases rapidly. At the approach of breakdown, the anode region of the PMT begins to emit light—some of which is detected by the photocathode.

According to some users, the measurement of low signal and dark current levels is affected by 1/f noise. Young (1969) concludes that reports prior to 1969, including his own, tend not to show this phenomenon, particularly where signal current is concerned. A detailed study by Coates (1972) concludes otherwise, citing a PMT with excess noise—that is one that generates a non-thermal component at high voltage, and hence at high gain (see Fig. 7.14). Excess counts, particularly at low signal levels, exhibit enhanced variance compared with the predictions of Poisson statistics. Various causes and cures have been suggested:

- voltage breakdown and the generation of light
- sharp discharge points within the structure
- leakage currents
- poorly regulated power supply
- unstable voltage divider (active dividers in particular)
- afterpulses and self-gettering
- housing design.

6.4 S/B in DC applications

Where a PMT is used for pulse counting or encoding, it is advisable to limit the gain to avoid the threat of excess dark counts, leading to breakdown. In these situations the required system gain must be realized by a combination of PMT gain and amplifier gain. The reverse applies to low-level current measurement: it is beneficial to operate the PMT at relatively high gain, where the relative leakage current contribution is small. This is already obvious from the information in Fig. 6.5. We can find the optimal operating region under DC conditions by plotting I_d/g versus g , as in Fig. 6.6.

The essence of the data presentation in Fig 6.6 lies in its simplicity, first recognized by Sharpe and Stanley (1962), who, like the present author, have recorded many such plots. The data required for Fig. 6.6 need not be absolute, thus inviting the use of relative gain on the abscissa; the goal is to uncover the voltage corresponding to the lowest point on the graph as the optimal operating point for that particular PMT. If the requirement is to select the best from a batch of PMTs, then it may be necessary to involve absolute gain. Optimizing the quality of performance in low-light-level applications has more to it than finding the preferred gain at which to operate.

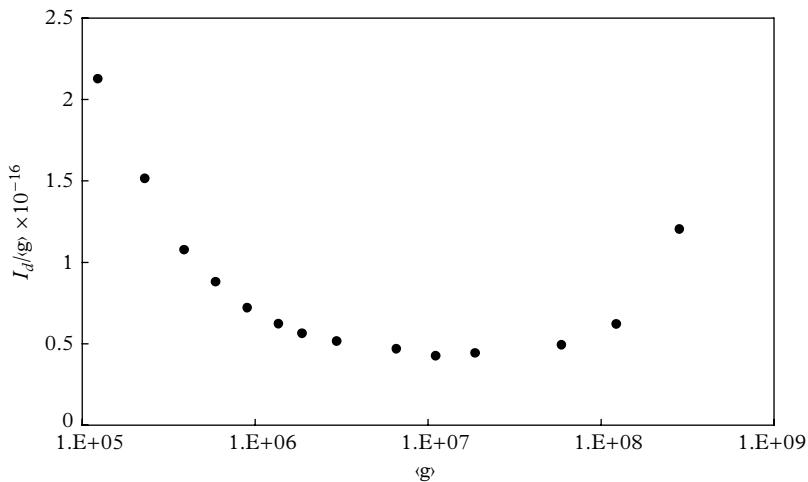


Fig. 6.6. The optimal operating gain for this PMT, when used for electrometer applications, is centred on $\langle g \rangle = 10^7$. The S/B ratio deteriorates markedly at $<10^6$ gain because of a leakage contribution, and also at ultra-high gain because of an increase in dark counts with HV. Note that the ordinate is an equivalent cathode current.

6

Photocathode sensitivity, S , plays an important role as does the optimal division of time allocated between measuring signal and background, discussed in 7.4.1.

6.5 Light generation within a PMT

An ideal PMT is one in which dark count remains constant as the voltage is varied. There are PMTs capable of providing this quality of performance at gains approaching 10^9 , but the majority of high-gain PMTs malfunction before this is achieved. The term breakdown refers to the generation of high anode current, sometimes exacerbated by electrical short circuits. Some devices recover while others may be left gassy. Light emission within a PMT, of sufficient intensity for the naked eye, forewarns of the imminent onset of breakdown and ultimately destruction. This was first reported by Krall (1967), who photographed the anode region of such PMTs. The present author, using an image intensifier, has observed that light emission occurs at moderate gains of $\sim 10^6$. The source of light has been ascribed to the impact of electrons on dynode surfaces and confirmed by the observation that the region around the last two or three dynodes, where current densities are highest, is most susceptible to light emission. Light generation is also evident from sharp metallic points and patches of alkali metals on the inside of the glass envelope. The contribution made by these light sources to dark counts depends on the degree of shuttering included in a particular device. PMTs

intended for photon counting invariably incorporate a constriction in the envelope against which a plate, usually part of the focussing arrangement, is pressed. The constriction serves to disrupt light piping down the envelope to the photocathode from sources that will now be discussed.

Light is generated by gamma rays, energetic electrons, muons, and other charged particles in the environment. Gamma rays interact in the glass of a PMT, producing electrons through three electromagnetic processes, and those electrons with sufficient energy produce light by Cerenkov emission. Photons produced in the window of a PMT are significantly more likely to be detected than those produced in other parts of the envelope (for the present discussion, we will take the envelope to refer to the glass enclosure, excluding the window). In any PMT, the window constitutes only one-third to one-eighth of the total volume of glass; consequently, radiation sources produce more light in the envelope than in the window. This can be understood in terms of the larger interaction mass of glass presented to incident gammas and, in the case of muons, it is the extra target area that is significant. A portion of light generated by Cerenkov emission propagates, via TIR, towards the window, where interaction with the photocathode may occur. This is an inefficient process but the proposed mechanism is easily established by temporarily attaching a beta source, such as ^{90}Sr , to a window; contributions to Regions A, B, and C will be apparent. To summarize: gamma interactions in the envelope contribute to A and B, and those which interact in the window contribute to C; charged particles passing through the envelope contribute to A, B, and C, and those passing through the window contribute to D.

PMTs are vacuum devices, but residual gas is always present within the envelope, regardless of the efficiency of the evacuation process. Internal surfaces retain atoms and molecules, some of which become detached by energetic secondary electrons. Those atoms that are ionized or excited emit light on returning to their ground states. Most PMTs incorporate a U-shaped metal shield, partially enclosing the back end of the multiplier, with an internal electrical connection to a middle dynode. This functions as a trap for ions produced in the region that escape the multiplier. A shield also serves to contain the light emission previously mentioned. The efficacy of an ion shield/light trap can be gauged by comparison with the performance of similar PMTs in which a shield is absent.

6.6 Sources of background

The basic processes responsible for background were well known 40 years ago, notably from the studies of Baicker (1960), Sharpe and Stanley (1962), Young (1966, 1969), and Krall (1967). Detailed studies on more than a single PMT have been reported by Barton *et al.* (1964), Gadsden (1965), Oliver and Pike (1968, 1970), Coates (1971, 1972a, b, 1973), and Wright (1983). Coates' work is particularly noteworthy, albeit restricted to the study of five samples. A series of articles appeared between 1971 and 1972 makes essential reading for those with a serious

interest in the physics of the subject. Coates initially made detailed studies of three RCA PMTs, all of which included the then recently introduced GaP, high-gain first dynode.

6.6.1 Region A: Undersized pulses

We can account for the origin of undersized pulses in the background, having explained in 5.5.3 how they arise in the signal spectrum. The presence of a low-energy subpeak is often missed, simply because it is unusual to operate at the gain required to reveal its presence. The observation that the ratio of the two peak positions in Fig. 6.7 is approximately the same as the gain of the first stage points to the origin of this small peak. A fraction of the light produced in the envelope and window by ionizing and Cerenkov radiation is collected by the first dynode, which consequently produces photoelectrons. Electrons are also generated at the first dynode because of its low work function. The pulse height distribution for background is represented by the solid dots in Fig. 6.7. Operation with the cathode open circuit is depicted by the triangles, but subtracting this distribution from the main one leaves an excess of small signals—the triangles. These are attributed to inelastic scattering, discussed in 5.5. Further confirmation concerning the origin of undersized pulses is given in Fig. 6.8.

6

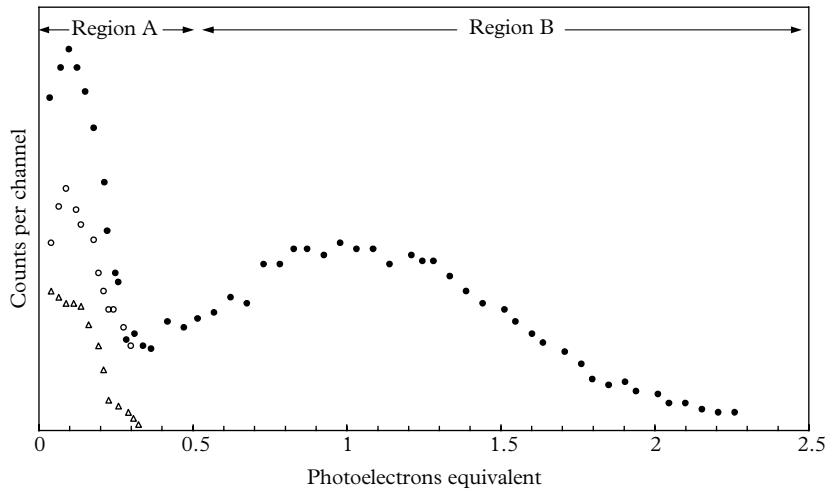
6.6.2 Regions B and C: Single- and multi-electron background

Dark counts in this region of the spectrum originate from the photocathode, either directly or indirectly. The direct component is attributed to thermionic emission and to photoemission arising from ionization and excitation processes within the window, while the sources of the indirect contribution lie elsewhere. The primary sources for indirect contributions produce light or ions, of which a portion interact with the photocathode, which in effect serves as a source of single electrons. It is generally assumed that thermionic emission is the major component of electrons in Region B. Strictly speaking, all one is entitled to say is that a fraction of the counts in Regions A and B are temperature sensitive, which is different from categorizing all of them as thermionic.

Multi-photoelectron background, C, derives from radioactivity in the ceramic and glass components, with a contribution from sources of environmental ionization. Afterpulses derived from Region B also contribute.

6.6.3 Dependence on temperature

The dependence of dark counts on temperature is illustrated in Fig. 6.9 for two 30 mm PMTs with bialkali photocathodes. It is clear that the experimental points do not follow the Richardson equation: there is an excess of counts between 25 and 45°C, and a deficit beyond 45°C. Coates (1972) also found a divergence between theory and experiment. The conclusion is that thermionic, or thermal, emission contributes to background but there are other sources that are also temperature dependent.



6

Fig. 6.7. The background spectrum (solid dots) has the same general shape as the one due to single photons, but features, such as the single-electron peak, are not as distinct. Open circles refer to operation with the cathode floating. Subtraction of the two distributions only accounts for a third of the small pulses, shown by the triangles.

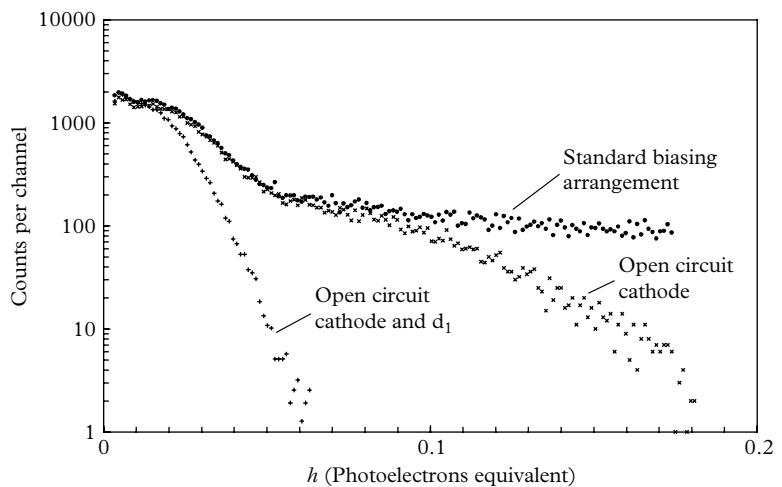


Fig. 6.8. Verification that the majority of small pulses in the background originate from dynodes. The scale on the abscissa has been expanded for this purpose. The biasing arrangement applies three times the inter-dynode voltage between the cathode and d_1 , resulting in a d_1 gain of 17; all other dynodes provide a gain of approximately 4.

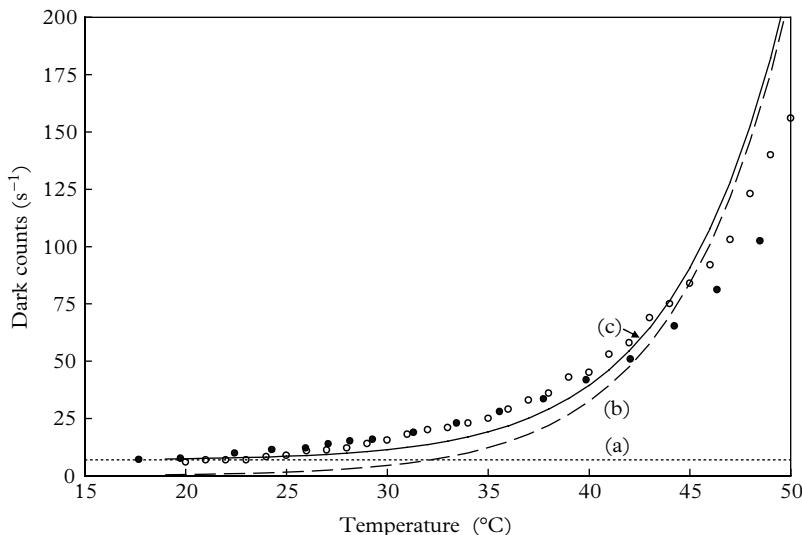


Fig. 6.9. Variation of background count with temperature. Curve (a) is the non-thermal component of the background, taken as 7 counts/s; curve (b) is calculated from the Richardson equation $N = AT^2 \exp(-eQ/kT)$, arbitrarily normalized and assuming a best-fit work function $eQ = 1.58$ eV. The solid curve, (c), represents (a) + (b). The open and closed circles refer to different bialkali PMTs of the same type.

6

6.6.4 Ionizing radiation

Gamma rays and electrons are sources of background counts in Regions B and C. They are present in a laboratory environment and in the envelope of PMTs. Viehmann *et al.* (1975) carried out extensive tests on a range of window materials suitable for space missions. Scintillation efficiency, ϵ , from which the list in Table 6.1 is derived, is the key parameter; it is the fraction of incident particle energy that is converted into visible light. Noting that plastic scintillator produces of the order of 5 to 10 photons/keV of energy loss, underlines the inefficiency of the process in window materials. The present author has performed tests using ^{57}Co , which emits a single gamma line of energy 122 keV. This is an ideal source for the present analysis because its energy is below the Cerenkov threshold of 175 keV—hence there is no Cerenkov contribution from ^{57}Co radiation (see 6.8.1). A small point source of ^{57}Co was chosen for the investigation, having first measured its emission rate over a 2π solid angle by placing it in contact with, and on the axis of, a cylindrical 2" NaI(Tl) crystal. Additional single-photoelectron counts were recorded, as indicated in the tail of the SER in Fig. 6.10, by placing the source in contact with a 4 mm thick borosilicate window of a 3" PMT. Substituting the PMT with a quartz window version shifted the distributions to the right in Fig. 6.11. The pulse height distribution produced by ^{57}Co is identical in shape to the SER in the two cases,

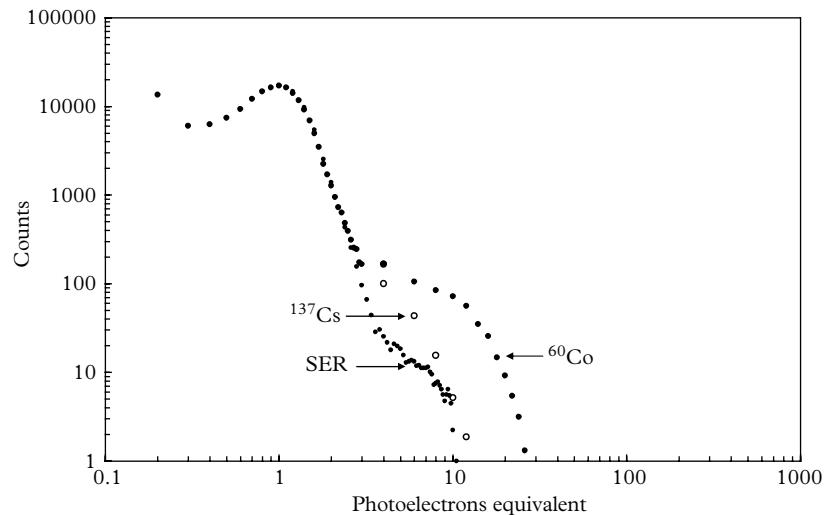


Fig. 6.10. The effect of placing gamma ray sources close to a PMT leads to a contribution in Regions A, B, and C of the spectrum. This 3" PMT has a 4 mm thick borosilicate window. Data marked SER represents the pulse height distribution for single-photon excitation alone.

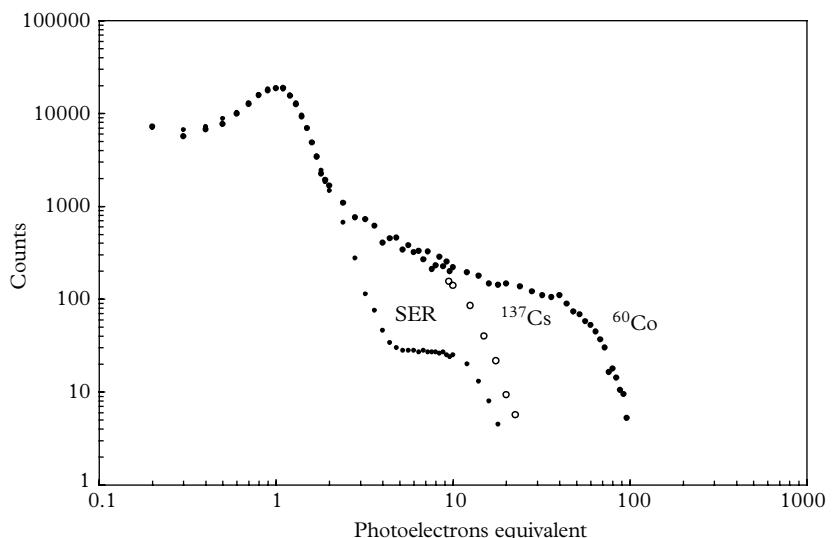


Fig. 6.11. Background spectra in a 3" PMT with a quartz window. The photon yield significantly exceeds that for the borosilicate PMT in the previous figure. Counts due to ^{57}Co are restricted to the SER region.

Table 6.1 The fluorescence yields of various materials measured using a ^{90}Sr source of electrons, due to Viehmann *et al.* (1975): second and third columns. The measurements in the fourth column are by the present author.

Material	Counts/MeV, $2\pi^a$	(%)	(%)
Plastic scintillator	250	3.00	—
7056 TM borosilicate glass ^b	2	0.02	0.023
9741 TM UV glass ^b	6	0.07	0.080
Suprasil TM (SiO_2) ^b	7	0.08	—
Spectrosil TM (SiO_2) ^b	7	0.08	0.01
Sapphire	20	0.24	0.14
MgF_2	8	0.10	—
MgF_2 (irradiated) ^c	25	0.30	—

^a The notation 2π refers to the measurement solid angle.

^b Corning Inc.

^c Irradiated with 2×10^{13} electrons.

whereas ^{137}Cs and ^{60}Co also contribute to Region C. The results of the present measurements are shown together with those of Viehmann *et al.* (1975) in Table 6.1.

It is concluded, for typical PMT window materials, that the contribution from ionization and excitation ranges from 2 to about 20 photons per MeV of energy deposited. The activity also depends on the history of the samples; there is considerable variation among samples from different manufacturers and there is even variation in batches from the same supplier. According to Viehmann *et al.* (1975), the ionization and excitation processes are associated with decay times ranging from minutes to many hours. However, the present author observed that light emission appeared to cease after less than one-tenth of a second after removal of the source of radiation, which might be explained by the very low doses that were applied.

6.7 Gamma background

There are three main sources of gamma radiation significant to PMT background. First, there are the ubiquitous naturally occurring radionuclides ^{40}K , ^{232}Th , and ^{238}U , found in the earth's crust. Radionuclides located inside the PMT, and in particular any within the window of a PMT, are of particular concern. Second, there are man-made sources at particle accelerators and commercial facilities contributing to a broad range of particles and gammas. Third, radiation belts surrounding the earth are of concern to satellite missions. The importance of all these sources is discussed in the sections that follow.

The contribution to PMT background from gammas of energy less than 175 keV derives from ionization and excitation only, since Cerenkov emission is not energetically possible (see Appendix B). The pulse height distribution is identical to the SER, which is readily verified with a ^{57}Co source of radiation, for example, as discussed in 6.6.4. Gamma rays are capable of producing light, in borosilicate glass, following Compton scattering and photoelectric capture. Light is also produced as a consequence of the Cerenkov effect, which is discussed fully in Appendix B. The magnitude of the contribution to background depends on the location of the source of gamma rays, its strength, and its energies. Sources may be found in the PMT environment, such as buildings and the ground upon which they are built, and even in the air (Radon). While massive shielding can be used to attenuate the contribution from such sources, those located within the PMT itself are particularly troublesome—especially if located within the window.

The contribution from a gamma ray of energy less than 2.62 MeV is easily determined using standard laboratory reference sources of a few microcuries in strength. The signals produced by ^{57}Co , ^{137}Cs , and ^{60}Co gamma rays fall within the same energy range as the natural background. They can be quantified by recording pulse height distributions for these sources when placed a few millimetres away from the window of a PMT under test. The count rate depends purely on the source strength, but what is important is establishing the region of the dark count pulse height distribution that is affected. Results are given in Figs 6.10 and 6.11 for a set of 3" type 9265B PMTs, each with a different window material of a standard 4.0 mm thickness.

The results in Fig. 6.11, for example, can be reconciled with those for MIPs, which produce 20 pe per millimetre of path length in quartz (see Appendix B). Considering the Compton effect only, the maximum electron energy that can be produced by a single interaction of a 1.33 MeV ^{60}Co gamma is 1.11 MeV. An electron of this energy has a range of 2 to 3 mm in glass (or quartz); it will produce the equivalent of 50 pe in quartz and about 20 pe in borosilicate glass. Electrons produced by the photoelectric effect carry the full gamma energy but the cross section for this process is negligible in glass, compared with the Compton contribution, so a peaked pulse height distribution should not be expected. The aforementioned radionuclides produce electrons, gammas, and alphas. The decay scheme for ^{40}K is comparatively straightforward compared with those for ^{232}Th and ^{238}U , which involve many daughter products. However, it is possible to relate the concentration of any isotope in parts per million (ppm), or parts per billion (ppb), to the total number of gammas emitted per kilogram per day. Detailed figures are given by McAlpine (2011), as follows: ^{40}K , 1 ppm = $285 \text{ d}^{-1} \text{ kg}^{-1}$; ^{232}Th , 1 ppb = $958 \text{ d}^{-1} \text{ kg}^{-1}$; and ^{238}U , 1 ppb = $2310 \text{ d}^{-1} \text{ kg}^{-1}$. Gamma rays, which are uncharged, produce light, but only indirectly as a consequence of an electromagnetic interaction in the envelope and window. The resulting electron, if sufficiently energetic, is the source of light by Cerenkov radiation, but only that portion of light that interacts with the photocathode produces signal. The electromagnetic processes involved here are the photoelectric effect, Compton scattering,

and pair production; in all these interactions, energy is transferred from the gamma ray to an electron in the window (or electrons, in the case of pair production). In the photoelectric effect and in pair production, the gamma ray imparts its entire energy in a single interaction, but in Compton scattering only a fraction of the energy is released. The surviving gamma ray, always of lower energy than the original, may undergo interactions in the glass, through any of the three processes mentioned, but noting that pair production is only possible for gammas of energy in excess of 2×0.511 MeV. The probability for the occurrence of these interactions can be expressed in various ways: where PMTs are concerned, the attenuation coefficient, σ , derived from cross-section formulae, applies.

The rate of background events is essentially proportional to the volume of the window because the contribution from the rest of the envelope is much reduced through solid angle and shielding considerations. The radiation environment in the immediate vicinity of a PMT is readily measured, in a qualitative way only, with a scintillation crystal, such as NaI(Tl). A typical spectrum for a crystal of dimensions $1\frac{1}{4}'' \times 1\frac{1}{2}''$ has the characteristic shape shown in Fig. 6.12, with identifiable peaks produced by the three naturally occurring isotopes. The details of a measured spectrum depend on the size of the crystal because there is a significant increase in high-energy detections with mass of the scintillator. Peak positions in Fig. 6.12 establish the energy scale. Note that there appear to be no discernible peaks beyond the thallium one (part of the uranium decay series) at 2.62 MeV. The upper spectrum in Fig. 6.12 refers to an unshielded crystal showing a uranium peak at 0.61 MeV, ^{40}K at 1.46 MeV, and the thorium peak at 2.62 MeV.

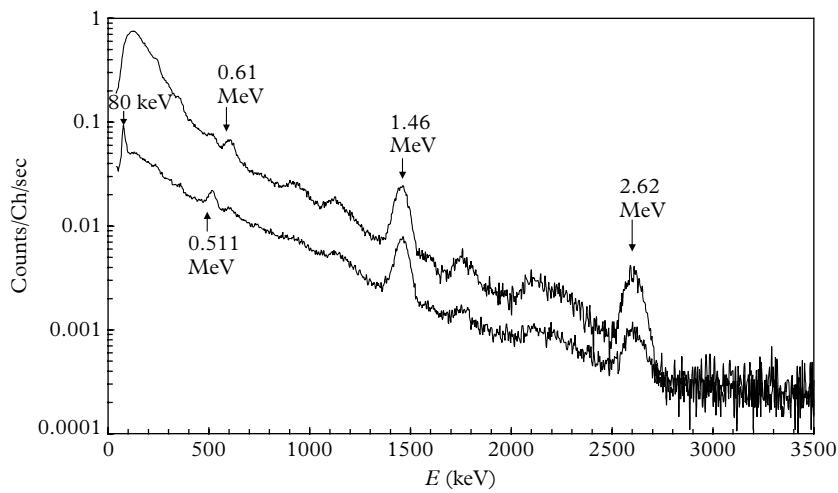


Fig. 6.12. The radiation background spectrum measured with a $1\frac{1}{4}'' \times 1\frac{1}{2}''$ NaI(Tl) crystal, unshielded (upper curve), and with lead shielding (lower curve). The peaks are characteristic of the radionuclides indicated.

The lower spectrum demonstrates the effect of a partially shielded crystal. The annihilation peak at 0.511 MeV shows quite prominently, indicating that the crystal itself has significant ^{40}K . The fluorescence peak at ~ 80 keV arises from the lead shielding.

6.7.1 Radionuclides within PMTs

PMTs are constructed of glass, ceramic, and metal. Since glass and ceramics are made from naturally occurring materials, we find that most of the activity emanates from such materials. Sand is the major constituent of glass, and some sources of this material are of lower activity than others. It was once a closely guarded secret but now all major PMT manufacturers know the sources of low-activity material. Potassium oxide is among the various additives used to give glass desirable physical qualities. Unfortunately, naturally occurring potassium contains 0.0118% of the long-lived isotope ^{40}K , and its omission from the constituents list affects what glass workers call cosmetic properties. The appearance of the glass is marked by striations, cords, and bubbles, none of which has any serious effects on the performance or longevity of the PMT. Yet, users express concern over such blemishes.

Synthetic quartz, known as fused silica, is a pure material made from silane gas (SiH_4). It is used in PMTs where either detection at UV wavelengths down to 160 nm is required, or freedom from radioactivity is essential. The radionuclide levels in this material are so low that it is difficult to quantify them with certainty. Absolute levels of radionuclides, listed in Table 6.2, are quoted in parts per million or parts per billion. These figures refer to activity levels in the window only; although other parts of the PMT contribute, the window is the major source. The metal parts are largely free of activity but the ceramic rods and side plates may contribute significantly if sufficiently impure. Ceramic material of 99.6% purity is available and it is similar in activity to B53 glass (McAlpine 2011). The total decays per minute, given in the last column, refer to the number of gamma rays of all energies within a band spanning 5 to 2640 keV, but note there is an accompanying spectrum of betas.

Table 6.2 Background levels in PMT windows made in various grades of borosilicate glass, compared with fused silica. Total decays per minute (DPM) is for a 50 mm diameter window of weight 30g.

Material	K (ppm)	Th (ppb)	U (ppb)	Total DPM
Standard	<60,000	<1000	<1000	<400
Low background	300	250	100	25
Ultra-low background (B53)	60	20	10	5
Fused silica	<5	<5	<5	<0.3

Table 6.3 The analysis of a special glass melt, Schott 8246, for the SNO and Borexino experiments. Measurements were made at several low-background facilities, located underground (McAlpine 2011). All entries refer to Schott 8246 except the last, which is for B53 glass (taken from McAlpine 2011).

Laboratory	K (ppm)	Error (ppb)	Th (ppb)	Error (ppb)	U (ppb)	Error (ppb)
California ¹	33	±3	18	±4	18	±2
Modane ²	30	±1	25	±8	23	±5
Guelph ³	—	—	15	±5	21	±3
Holborn ⁴	33	±8	10	±8	34	±3
Gran Sasso ⁵	30	—	20	—	25	—
Holborn ⁴ (B53)	60	±15	30	±10	30	±20

¹ Caltech, X X Hutamian.

² Laboratoire Souterrain de Modane, CENBG, France, Ph Hubert.

³ University of Guelph, Ontario Canada, JJ Simpson.

⁴ Birkbeck College, London, J C Barton.

⁵ INFN, l'Aquila, Italy, C Arpsella.

An interesting comparative study was carried out in the early 1990s, pursued mainly by the SNO and Borexino collaborations involving massive kiloton detectors located deep underground. The rarity of the scientific events demands minimal radionuclides throughout the detector. Any such sources located within the glass of the PMTs, as previously discussed, are particularly undesirable. The investigators in these two experiments ordered a special melt of Schott 8246 glass, possibly the lowest activity glass ever produced. Laboratories listed in Table 6.3 measured the activity of this glass and also special glass provided by Electron Tubes Ltd and designated B53. Measurements were performed over weeks or even months using Ge detectors, but the scatter in the results highlights the difficulties in making precise measurements of materials with very-low-activity levels.

A requirement for an all-quartz PMT is periodically put to manufacturers. The activity of such a PMT would be measured in terms of the emission of a few gammas per hour. Unfortunately, it is not possible to bring out the pin connections through quartz because of the significant difference in expansion coefficients for metals and for fused silica. This can be overcome by the use of so-called graded seals consisting typically of a set of five fused glass rings. Each ring has a slightly different expansion coefficient from its neighbour, thus providing the means to couple the quartz section of the envelope to the base of the PMT, illustrated in Fig. 6.13. Although the seals are thin, they are still too active for the most critical whole body arrangements, and for massive neutrino detectors.

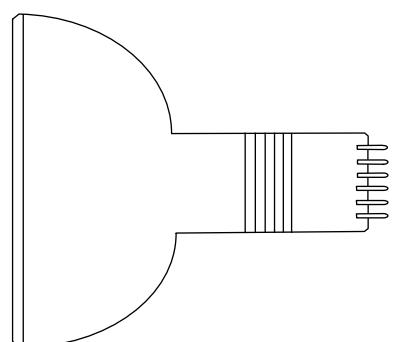


Fig. 6.13. A 5" quartz PMT showing the location of graded seals on the envelope.

Typical activities for graded seals are 7200 ppm, 1030 ppb, and 420 ppb, for K, Th, and U, respectively. Hamamatsu Photonics offers a very-low-activity PMT comprising a fused silica window and a metal envelope, whereas Electron Tubes Ltd developed an all-quartz PMT with a band of molybdenum close to the pin connections. This type of PMT is rare and very expensive.

6.8 Cosmic ray muons

Table 6.4 The composition of the cosmic ray flux at sea level. Entries are approximate and depend on the assumed energy cut-off. The muon flux at sea level is easily remembered as, approximately, 1 muon/cm² min⁻¹. The angular distribution of muons at sea level is given by $I_0 \approx I_\theta \cos^2 \theta$.

Charged particle	Composition (%)
Muons	63
Neutrons	21
Electrons	15
Protons/pions	<1

Contributions to Region D are particularly conspicuous in all PMTs of diameter greater than 50 mm. The rate of events is proportional to the area, and the size of the pulses is proportional to the window thickness. Sometimes users mistakenly conclude that these large pulses signify that the PMT is breaking down. The presence of cosmic rays and their local rate are easily verified with a NaI(Tl) crystal assembly. Extending the energy scale to 40 MeV uncovers a rather featureless spectrum with a broad peak at 15 MeV, shown in Fig. 6.14. The events in the region spanning 5 to 40 MeV are derived from the combined contributions from the hard and soft components of the cosmic ray background—the peak at 15 MeV is caused by muons (see Table 6.4). This MIP suffers an energy loss in NaI(Tl) of about 1.5 MeV/g or, equivalently, 5 MeV per centimetre of path length. Since the typical track length in the crystal is of the same order as its dimensions, that is, about 3 cm, the peak at 15 MeV is compatible with the stated energy loss.

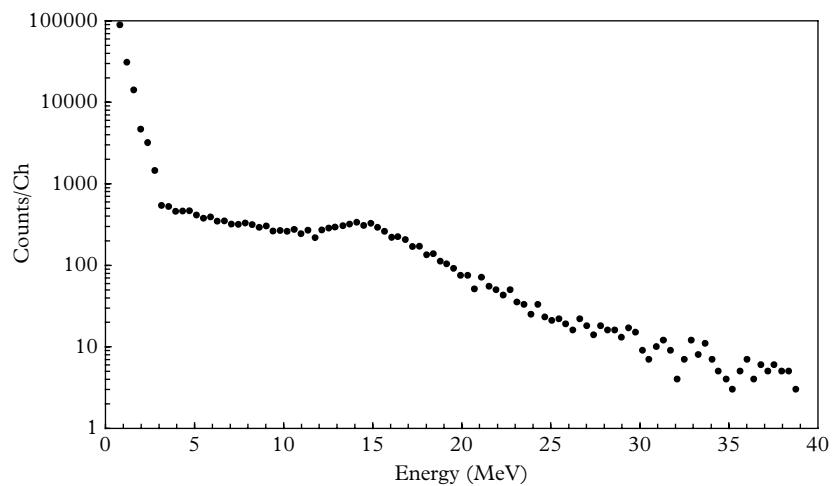


Fig. 6.14. The background spectrum beyond 5 MeV, measured with a NaI(Tl) assembly. The energy distribution is primarily due to the soft and hard components of the cosmic ray background.

6.8.1 Cerenkov emission in a PMT window

Pulse height distributions are shown in Fig. 6.15 for three PMTs, all of diameter 76 mm, but fabricated from different window materials. For clarity of presentation, only the high-energy part of each spectrum is shown. The ratio of the peak positions of borosilicate, UV transmitting glass, and Spectrosil is 1:1.83:3.17. This ratio compares favourably with that derived from theoretical calculations and given in Fig. B.5 (Appendix B), that is, 1:2.35:4.2. The windows are all 4 mm thick, and the peak position is expected at 112 pe in the PMT with a quartz window and yet it occurs at nearly double the pulse height. The explanation for this apparent discrepancy lies in optical enhancement due to TIR, as discussed in 3.13. The critical angle for glass is 42° whereas the Cerenkov angle is 48° and, consequently, all light produced by a vertical muon in traversing a PMT window, held horizontally, is trapped by TIR. The Cerenkov light thus has multiple opportunities to react with the photocathode, leading to an enhanced QE. In predicting the photoelectron yield for the Cerenkov effect (see Appendix B, Fig. B.5), no allowance was made for this enhancement, and the assumed QEs were those measured in the usual way.

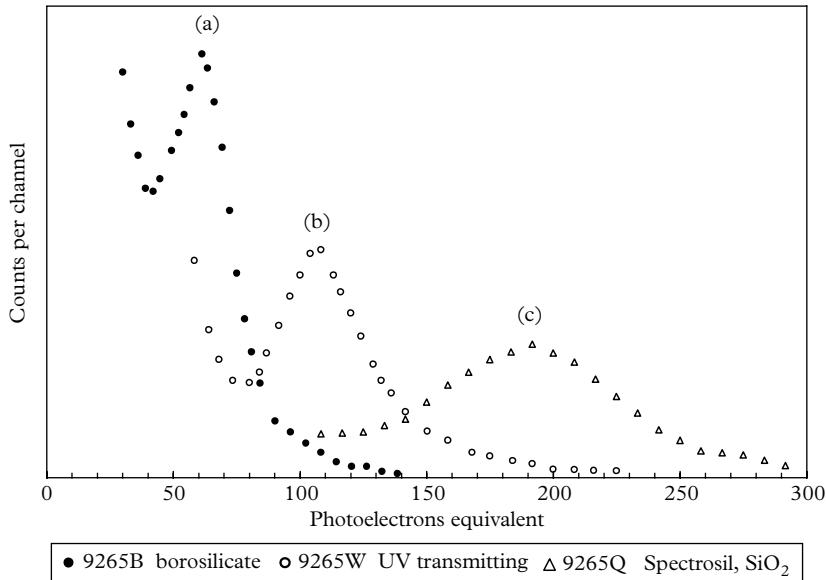
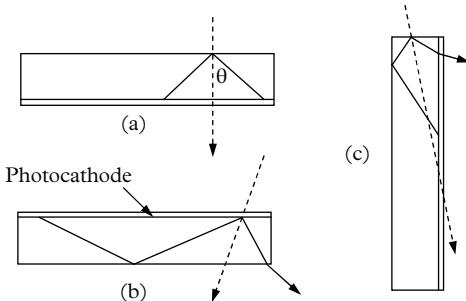


Fig. 6.15. Differential pulse height distributions for cosmic ray muons, showing the effect of different window materials. Windows are located horizontally and upwards-facing. They are 4 mm thick and hence the yields in photoelectrons per millimetre of path length for any bialkali PMT are of the order 15, 26, and 50 for (a), (b), and (c), respectively.

6.8.2 PMT orientation

The counts in Regions A, B, and C of the background spectra are independent of PMT orientation. This does not apply to the most energetic pulses in the background spectrum, Region D. There are three considerations: first, the muon angular distribution rolls off as $\sim \cos^2\theta$ so the acceptance aperture will be a maximum when the plane of the window is horizontal; second, the longest muon path lengths



6

Fig. 6.16. A relativistic muon, shown by a dashed line, produces light in proportion to its track length in the window. All light generated by a vertical muon is trapped through TIR in (a). A proportion of the light escapes in (b). The orientation shown in (c) produces the occasional large pulse.

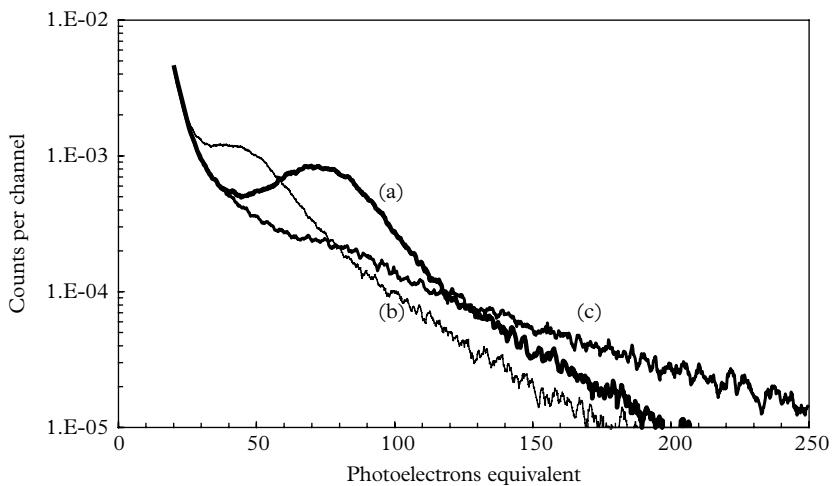


Fig. 6.17. The effect of orientation on the high-energy region of the background. The pulse height distributions are for a 5" 9390B with a diameter of 134 mm and a borosilicate window of thickness 6.4 mm. The window orientations are (a) horizontal, facing up; (b) horizontal, facing down; and (c) vertical.

occur with the plane of the window vertical and hence the presence of a long tail, extending out to 500 pe, is evident in Fig. 6.17(c); third, more light is intercepted by the photocathode when the PMT is horizontal and facing upwards. Some 25% more light is collected in this orientation, compared with face down, as can be seen by comparing Fig. 6.17(a) and (b). Furthermore, resolution is noticeably better in Fig. 6.17(a) compared with (b). Observations shown in Fig. 6.17 are readily explained in terms of the geometrical arrangements outlined in Fig. 6.16.

6.9 Means for reducing background

Low-background counting systems, which may include a 4π anti-coincidence shield, are capable of vetoing muon events with high efficiency. In any case, the signals produced where scintillators are employed are often outside the energy range of interest and may thus be eliminated on that basis. Contributions from naturally occurring radioisotopes may be reduced by the use of shielding with low-activity lead and copper. The materials located within the shielding need careful scrutiny and selection—especially the PMT.

The number of dark counts and the dark current, in all of Regions A, B, C, and D, varies in proportion to the area of the PMT window. In D, the size of the pulses from muons is dependent on the window thickness; the rate is proportional to the square of the diameter. For gamma radiation, from a source internal or external to the window, the rate of events is proportional to the volume of the window, but the size of the background pulses is independent of dimensions. The thickness, and hence volume, of a flat 2" window can be reduced to ~ 0.5 mm, provided that it is sufficiently domed to withstand atmospheric pressure. Ten- and twelve-inch hemispherical PMTs satisfy the mechanical requirements with a 2 to 3 mm thick envelope, taking advantage of the inherent strength of such an assembly.

6.10 Summary

The contributions to PMT background are as follows:

- photocathode type
- area of the photocathode
- types of dynode and their number

- window material
- selection performed by a manufacturer
- gain, applied voltage and its polarity
- ageing under applied HV
- recent light exposure under operation
- details of the voltage divider design
- the nature of the enclosure or housing
- temperature of operation
- the nuclear radiation environment.

REFERENCES

- Baicker, J. A. (1960). Dark current in photomultipliers. *IRE Trans. NS*, **7**, 74–80.
- Barton, J. C., Barnaby, C. F., and Jasani, B. M. (1964). An investigation of noise in Venetian blind photomultipliers. *J. Sci. Instr.*, **41**, 599–604.
- Coates, P. B. (1971). Noise sources in the C31000D photomultiplier. *J. Phys. E: Sci. Instr.*, **4**, 201–7.
- Coates, P. B. (1972a). Thermionic emission from photocathodes. *J. Phys. D: Appl. Phys.*, **5**, 1489–98.
- Coates, P. B. (1972b). Photomultiplier noise statistics. *J. Phys. D: Appl. Phys.*, **5**, 915–30.
- Coates, P. B. (1973). A theory of afterpulse formation. *J. Phys. D: Appl. Phys.*, **6**, 1862–9.
- Gadsden, M. (1965). Some statistical properties of pulses from photomultipliers. *Appl. Opt.*, **4**, 1446–52.
- Krall, H. R. (1967). Extraneous light emission from photomultipliers. *IEEE Trans. NS*, **14**, 455–9.
- McAlpine, R. M. (2011). Photomultipliers for low background applications. ET Enterprises Reprint Series ET Enterprises Reprint Series, RP092 . ET Enterprises Ltd, Uxbridge. Available at http://www.et-enterprises.com/files/file/technical-information/rp092_pmts%20for%20low%20background%20appl.pdf (accessed 29 Jan 2017).
- Oliver, C. J. and Pike, E. R. (1968). Measurement of low light flux by photon counting. *J. Phys. D: Appl. Phys.*, **1**, 1459–68.
- Oliver, C. J. and Pike, E.R. (1970). The edge effect in multiplier statistics. *J. Phys. D Appl. Phys.*, **3**, L73–L74.
- Sharpe, J. and Stanley, V. (1962). ‘Photomultipliers for tritium counting’. In *Tritium in the Physical and Biological Sciences, Vol. 1*. International Atomic Energy Agency, Vienna, 211–29.

- Viehmann, W. *et al.* (1975). Photomultiplier window materials under electron irradiation: Fluorescence and phosphorescence. *Appl. Opt.*, **14**, 2104–15.
- Wright, A. G. (1983). An investigation of photomultiplier background. *J Phys E: Sci. Instr.*, **16**, 300–7.
- Young, A. T. (1966). Cosmic ray induced dark current in photomultipliers. *Rev. Sci. Instr.*, **37**, 1472–81.
- Young, A. T. (1969). Photometric error analysis. IX. Optimum use of photomultipliers. *Appl. Opt.*, **8**, 2431–47.

7

Measurement of low
light flux

7.1 The physical nature of light detection

Photoemission is a quantum mechanical process that produces an output in the form of a train of random events. Periods of high light intensity are characterized by correspondingly high photoemission rates; conversely, periods of low intensity relate to low photoemission rates. The statistical treatment of photoemission assumes that the time averaged intensity of the light source is constant over repeated observation intervals, T . This assumption is valid for most sources of practical interest, such as lasers, LEDs, thermal devices, scintillators, thermoluminescence, and bioluminescence. The field of quantum optics is abstruse for the non-specialist, with notable articles relevant to PMTs appearing in the late 1950s, an example of which is the key paper by Mandel (1959). Considerable literature since 1959 may be traced starting with the contributions from Saleh (1973), Oliver and Pike (1968), and their many co-workers (see e.g. Foord *et al.* 1969). The important result, and the basis of this chapter, is that the probability of receiving n photons within time T is given by the Poisson distribution

$$P_{\mu}(n, T) = \exp(-\mu T) \times (\mu T)^n / n!, \quad (7.1)$$

where the mean rate of photons is μ per second, and constant. Statistically speaking, such systems are said to be stationary.

7

7.1.1 Noise and background

The statistical processes involved in photodetection are such that a repeated measurement under identical conditions is unlikely to produce the same number of counts, because of ‘noise in signal’. This noisiness is illustrated in Fig. 7.1(a), with the inescapable observation that any measured mean detection rate of photoelectrons depends on the location of the time band over which the counting takes place: moving this band to the left or to the right on the time axis leads to a different estimator of the mean rate.

It is important to understand that ‘noise’ is not synonymous with ‘background’; the latter refers to either dark counts or dark current. Some constituents of PMT background are more appropriately described as ‘unwanted signal’. The technical literature abounds with this confusion in terminology, and the reader needs to be on guard to decide whether reference is being made to a statistical quantity, noise, or to a background signal. The PMT background rate B is itself noisy and therefore embodies noise in background, in the same way that signal shows noise in signal. The aim is to examine the effect of this noise, in the various detection methods, ultimately for the purpose of selecting the best one for a given application.

7.2 Measurement modes

There are three experimental methods for quantifying the flux of light incident on a photodetector:

- counting output pulses initiated by photoelectrons—known as photon counting
- measuring the DC current flowing at the anode—referred to as analogue detection, or charge integration
- measuring the rms noise in the anode current—described as shot noise power detection

A range of sophisticated instrumentation is available to the user, although the basic detector systems described in this chapter may prove adequate. Of the three detection methods, counting the number of pulses initiated by single photoelectrons offers the most direct method for quantifying light flux. In some instances, information relating to photon arrival times is required, and photon counting is the only technique that can be adapted for this purpose. A discriminator and a scaler are fundamental equipment for photon counting. Analogue detection may be realized with a multimeter or an electrometer, while shot power measurements may be made with a true rms meter.

7

7.2.1 Photon counting

Detection and measuring schematics are illustrated in Fig. 7.1 for photon counting. Row (a) represents the arrival time of photons, 25% of which convert to photoelectrons. This is followed by multiplier amplification shown in (b). The variability in pulse height is a consequence of multiplier noise, but this contribution can be removed by including a discriminator with a set acceptance threshold. Discriminator output is shown in (c). In this example, the last seven pulses in sequence (a) only produce one photoelectron, shown in (b) but the output signal is below threshold and is therefore unproductive. The second pulse in the closely spaced pair falls within the dead time of the discriminator and does not contribute. The remaining pulses generate the set of digital pulses recorded in (c). The time axis is scaled in arbitrary units: for example, the abscissa is typically expressed in units of microseconds or nanoseconds.

7.2.2 DC detection

A schematic describing analogue measurement is given in Fig. 7.2. The PMT output is connected to a parallel combination of R and C, with the choice of time constant for DC detection largely dictated by the time structure of the light signal.

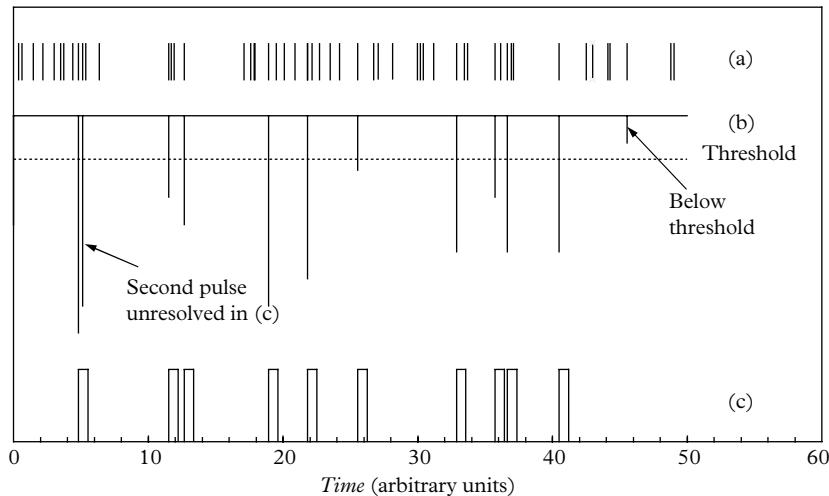


Fig. 7.1. Random arrival times of photons are illustrated in (a); a portion, η , initiates single-photoelectron signals of variable height at the anode, shown in (b). Standard digital pulses, derived from signals that exceed a fixed threshold, are shown in (c). Note the short time constant of the anode pulses must be preserved where high count rates are expected.

7

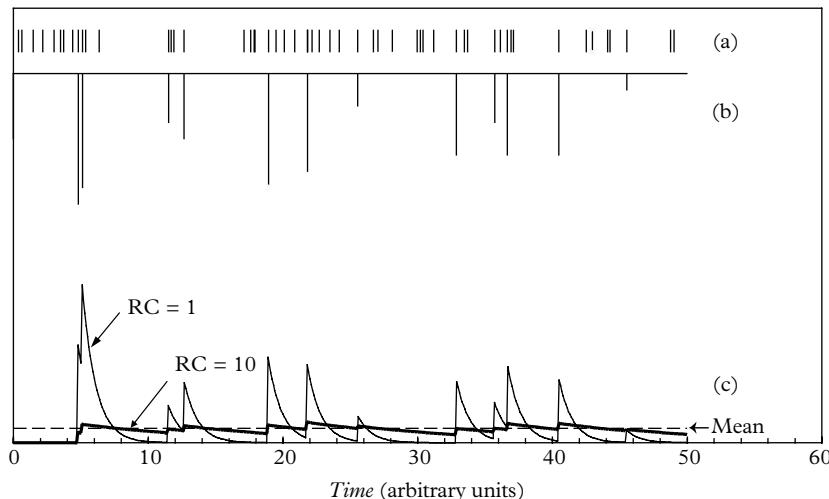


Fig. 7.2. An example of analogue detection based on the same sequence of photons adopted in the previous figure. The beneficial effect of smoothing is obvious in (c). Output signals have been inverted for presentation purposes.

If it is slowly varying, a relatively long time constant is appropriate; the output signal so generated is reasonably smooth with a tendency to follow the light input. The faint representation refers to a unity time constant, $RC = 1$, and the heavy line to $RC = 10$, achieved by increasing the capacitance—the dashed line is the mean. Analogue detection involves integrating output charge pulses, by means of a low-pass filter, to provide a DC level proportional to the light flux. Note that R determines the mean output voltage, while C produces smoothing.

There is a range of nuclear radiation monitors that generate a DC output voltage proportional to the activity of the source: this can be achieved in the manner of Fig. 7.2. An alternative is to first digitize pulses, as in Fig. 7.1, followed by integration. This effectively eliminates multiplier noise, resulting in reduced display flicker (further considered is provided in 7.5.10).

7.2.3 Shot noise power detection

The third technique for making low-level-light measurements is more ingenious than practical. Noise power detection is based on the shot noise formula, applied to the statistics of photoemission and multiplier noise. Application of this technique was first reported by Pao and Griffiths (1967). There have been few subsequent references to its use for low-light-level measurement, although the technique finds favour in linearity assessment of CCDs used in astronomical detectors (Janesick *et al.* 1987). This is discussed in 9.4.3.

The shot noise formula, due to Schottky, is

$$\langle i_k^2 \rangle^{1/2} = (2eI_k\Delta f)^{1/2}, \quad (7.2)$$

where $\langle i_k^2 \rangle^{1/2}$ is the rms noise in the mean cathode current, I_k , and Δf is the noise bandwidth. The signal-to-noise ratio is

$$S/N = I_k / (2eI_k\Delta f)^{1/2} = I_k^{1/2} / (2e\Delta f)^{1/2}, \quad (7.3)$$

which can be determined in terms of voltage, given a true rms meter:

$$V_{DC}/V_{rms} = \left(\frac{I_k}{2e\Delta f} \right)^{1/2}. \quad (7.4)$$

Therefore, I_k is proportional to $(V_{DC}/V_{rms})^2$ in a power relationship. Equations (7.2) to (7.4) apply to a vacuum photodiode but adaption to a PMT is easily done. Assuming an ideal PMT with collection efficiency F , we can apply (7.2) to the anode by replacing $\langle i_k^2 \rangle^{1/2}$ by $\langle i_a^2 \rangle^{1/2}/G$, and I_k by I_a/G where $\langle G \rangle$ is the PMT gain; by definition, $\langle G \rangle = I_a/I_k$:

$$\langle i_a^2 \rangle^{1/2} = (2e\langle G \rangle\Delta f)^{1/2} I_a^{1/2}. \quad (7.5)$$

Even if unknown, $\langle G \rangle$ and Δf are constant, and (7.5) leads to a straight-line relationship between $\langle i_a^2 \rangle^{1/2}$ and $I_a^{1/2}$ of slope $(2e\langle G \rangle\Delta f)^{1/2}$. Equation (7.5) can

be extended by replacing $\langle G \rangle$, which is difficult to measure, by $F\langle g \rangle$. A term representing ENF of the multiplier completes the formula for a PMT:

$$V_{DC}/V_{rms} = (2e\Delta f F\langle g \rangle I_a)^{1/2} \left(1 + \text{var}(g)/\langle g \rangle^2\right)^{-1/2}. \quad (7.6)$$

Clearly, if ENF, $\langle g \rangle$, and Δf are known, then (7.6) provides a means for determining F . This is not of immediate concern, whereas confirmation of the linear relationship between V_{DC}/V_{rms} and $I_a^{1/2}$ is. Note that e , Δf , F , $\langle g \rangle$, and $1 + \text{var}(g)/\langle g \rangle^2$ are all constants.

Experimental results taken with an inexpensive 3½-digit multimeter, (type Amprobe 37XR-A), for which the manufacturer quotes a bandwidth of 1.955 kHz, are shown in Fig. 7.3. A linear relationship is clearly followed, up to an anode DC current of $\sim 10 \mu\text{A}$, but beyond this level measurements are increasingly affected by gain change induced by the voltage divider, covered in 13.3. Although the shot noise power method offers no advantages over straight DC detection, it provides a means for measuring DC linearity, favoured by CCD users (Janesick *et al.* 1987).

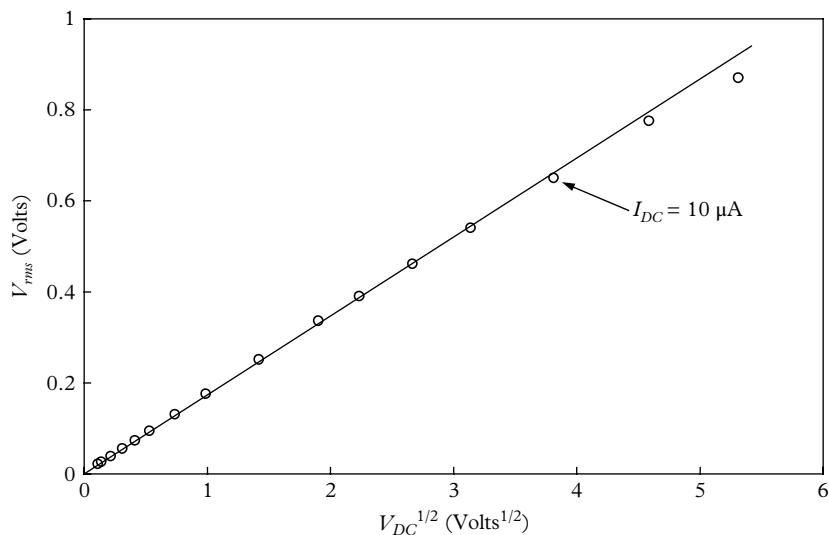


Fig. 7.3. Illustrating the shot noise relationship between noise voltage, V_{rms} , and V_{DC} . Dynamic range may be extended by adopting an active voltage divider. The anode circuit consists of $1\text{M}\Omega \parallel 50\text{pF}$. Note the deviation from linearity, due to an all-resistor voltage divider, is evident for I_{DC} beyond $\sim 10 \mu\text{A}$.

7.3 Detection methods and weighting

The inescapable statistical nature of light detection manifests itself as noise in signal. Single-photon excitation, assumed for this section, produces anode pulses with a range of pulse heights, h . There is a choice of three detection methods, and the aim is to compare the statistical performance of each. We follow the approach of Young (1969), by categorizing the different modes of PMT operation in terms of weighting functions, $w(h)$. An ideal quantum detector has the signal-to-noise ratio

$$S/N = 1/(n\eta)^{1/2} \quad (7.7)$$

where n is the number of incident photons per unit time, and η is the QE. Equation (7.7) refers to photoelectron noise only and applies equally to all detection schemes. The immediate interest lies in the multiplication process for the chosen detection method.

The output from a PMT consists of charge pulses quantified by the differential pulse height distributions $s(h)$ and $b(h)$, for signal and background counts, respectively.

In pulse counting, only those anode pulses with amplitude h above a set threshold or between an upper and a lower threshold are counted. Hence, the contributions to signal are weighted by a factor, $w(h)$, which is equal to unity, and performance is said to be digital.

In anode current detection, also known as charge integration, each pulse is weighted by its height or charge, and $w(h) = h$. Performance is described as analogue but is generally referred to as DC detection.

In the shot noise power method, the mean square of the instantaneous current is measured. Thus, each pulse is weighted by $w(h) = h^2$.

The three weighting functions are defined in Fig. 7.4.

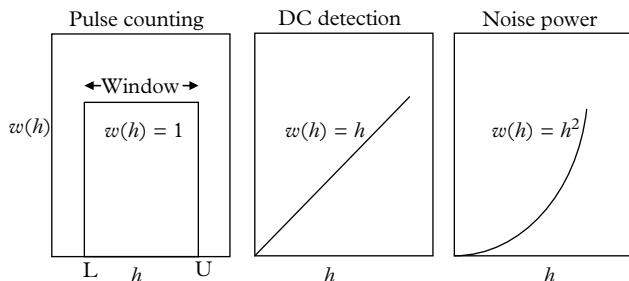


Fig. 7.4. Weighting functions, $w(h)$, for pulse counting, DC detection, and shot power measurement. It is common practice in pulse counting to omit the upper threshold, U , by counting all pulses $>L$.

Direct measurement of $s(h)$ is not possible: instead we always measure a composite distribution, $p(h)$, that includes a background contribution, $b(h)$:

$$p(h) = s(h) + b(h).$$

The signal distribution is deduced by subtracting $b(h)$. But it is assumed in the present analysis that the background count rate is negligible and hence it is ignored in the treatment that follows; S is defined as the measured quantity associated with the particular technique adopted and it is characterized by a weighting function, $w(h)$, with

$$S = \int_0^\infty w(h) s(h) dh. \quad (7.8)$$

Considering unit time and Poisson statistics, if there are on average $s(h)dh$ pulses within h and $(h + dh)$, the variance in this quantity is $\sigma_s^2 = s(h)dh$ and, in the measured quantity, S , it is

$$\sigma_S^2 = \int_0^\infty \left[\frac{\partial S}{\partial s(h)} \right]^2 \sigma_s^2 = \int_0^\infty \left(w(h) \right)^2 s(h) dh. \quad (7.9)$$

Note the differentiated function within square brackets is the standard transform used in error propagation (see e.g. Bevington and Robinson 1992). Young introduces a ‘degradation’ factor, Δ , to serve as a measure of the signal-to-noise performance for each scheme.

7

7.3.1 Detection techniques

In photon counting, only those pulses are registered that fall within a window defined by a lower threshold, L , and an upper threshold, U . Pulses that satisfy these conditions are counted with unity weight, $w(h) = 1$, and all others carry zero weight. The multiplier produces a spread in pulse heights at the anode but this noise is eliminated by the standardizing action of the window discriminator. Hence the signal-to-noise ratio for pulse counting, Δ_{pc} , is unity.

In DC detection, each anode pulse contributes to S in proportion to its height, and hence $w(h) = h$. The functions S and σ_s^2 are given by the moments μ_1 and μ_2 :

$$\begin{aligned} S &= \int_0^\infty h \times s(h) dh = \mu_1, \\ \sigma_S^2 &= \int_0^\infty h^2 \times s(h) dh = \mu_2. \end{aligned} \quad (7.10)$$

The signal-to-noise ratio is $(\mu_1^2/\mu_2)^{1/2}$, and the moments are given for a Poisson distribution in 4.10.4 as $\mu_1 = \delta$, and $\mu_2 = \delta(1 + \delta)$, where δ is the stage gain,

assuming that only the first stage contributes to noise. It is sufficient for present purposes to consider the first dynode as the only source of multiplier noise—this is justified by the analysis in 4.10.4 and Fig. 4.8. Therefore, for Poisson statistics, based on $\delta = 4$,

$$\Delta_{DC} = \mu_1^2 / \mu_2 = \frac{\delta^2}{\delta + \delta^2} = \frac{\delta}{\delta + 1} = 4/5. \quad (7.11)$$

Note that Δ_{DC} is the reciprocal of the relative variance term $1 + \text{var}(g)/g^2 = (\delta + 1)/\delta$, which was derived in 4.13.

In the shot noise method, the mean square of the instantaneous current is measured so that each pulse is weighted by the square of its height, $w(h) = h^2$:

$$\Delta_{sn} = \mu_2^2 / \mu_4 = \frac{[\delta(1 + \delta)]^2}{\delta^4 + 6\delta^3 + 7\delta^2 + \delta} = 400/756 = 0.53.$$

The treatment so far is based on the assumption of Poisson statistics for the multiplier; it represents the best achievable performance but the extreme case of an exponential SER, which simulates a poor-quality PMT, is also worthy of investigation. The moments for a continuous pulse height distribution, $\exp(-h)$, can be calculated as

$$\mu_n = \int_0^\infty h^n \exp(-h) dh = n!,$$

and hence

$$\begin{aligned} \Delta_{pc} &= 1, \\ \Delta_{DC} &= \frac{\mu_1^2}{\mu_2} = 1/2, \\ \Delta_{sn} &= \frac{\mu_2^2}{\mu_4} = 4/4! = 1/6. \end{aligned}$$

Results are collected in Table 7.1 for ease of comparison.

In practice, Δ_{pc} is always less than unity because of electronic noise (jitter) on signals close to the set threshold, although negligible for a quality discriminator. The entries in Table 7.1 reveal the importance of an SER with low dispersion. The rare use of the shot noise method in low-light-level detection is explained by its

Table 7.1 Figure of merit for three detection modes. Signal-to-noise ratios are given by the square root of the quoted delta values.

SER	Δ_{pc}	Δ_{DC}	Δ_{sn}
Poisson	1.00	0.80	0.53
Exponential	1.00	0.50	0.17

poor signal-to-noise ratio. The analysis adopted in this section brings out the salient points concerning the modes of detection, but without consideration of dark counts and the optimal allocation of measurement time between $p(h)$ and $b(h)$. Young (1969) treats this topic in great detail.

7.4 Difference between two count rates

It is usual in low-level counting applications to involve subtraction to arrive at the source strength. The required quantity is signal minus background, treated statistically. In the present context, we have: R_1 is the rate of single-photoelectron events, comprising signal plus background, and R_2 is background alone. Subtraction leads to an estimate of the signal count rate, $R = R_1 - R_2$, but calculating the variance is not so obvious. If we assume n_1 counts are recorded in time t_1 , and n_2 in time t_2 , the count rates and their standard deviations are

$$R_1 = \frac{n_1}{t_1} \pm \frac{n_1^{1/2}}{t_1}$$

$$R_2 = \frac{n_2}{t_2} \pm \frac{n_2^{1/2}}{t_2}.$$

The difference in rates is

$$\begin{aligned} R &= \frac{n_1}{t_1} - \frac{n_2}{t_2} \pm \left(\frac{n_1}{t_1^2} + \frac{n_2}{t_2^2} \right)^{1/2} \\ &= R_1 - R_2 \pm \left(\frac{R_1}{t_1} + \frac{R_2}{t_2} \right)^{1/2}. \end{aligned} \quad (7.12)$$

7

7.4.1 Efficient time allocation

Suppose that a fixed time, $T = t_1 + t_2$ is available for determining R , we wish to know the optimal allocation of T between t_1 and t_2 that minimizes the uncertainty

$$\pm (R_1/t_1 + R_2/t_2)^{1/2}.$$

Noting that T is a constant and $\partial/\partial t_1 = -\partial/\partial t_2$

$$\frac{\partial}{\partial t_1} \left(\frac{R_1}{t_1} + \frac{R_2}{t_2} \right)^{1/2} = \frac{1}{2} \left(\frac{R_1}{t_1} + \frac{R_2}{t_2} \right)^{-1/2} \left(\frac{-R_1}{t_1^2} + \frac{R_2}{t_2^2} \right). \quad (7.13)$$

The uncertainty in R is minimal when the term

$\left(\frac{-R_1}{t_1^2} + \frac{R_2}{t_2^2} \right)$ is zero, from which it follows that

$$t_1/t_2 = (R_1/R_2)^{1/2}. \quad (7.14)$$

The implications of the relationship in (7.14) are perhaps unexpected, but important: more time should be spent on the measurement of the higher rate

with times chosen proportional to the square roots of the rates R_1 and R_2 . Signal recovery of a weak intensity from a much more intense one is a special case often encountered in photon counting. When the signal counts are a small fraction of those contributed by background, the time available should be split equally. The theory given here also applies to low-background counting involving radioisotopes.

Let R_1 represent the rate of the signal, S , when measured in the presence of background, B ; R_2 is the count rate in the absence of signal. A feeble signal implies R_1 and R_2 are similar in magnitude and thus, from (7.14),

$$t_1 = t_2 = \frac{1}{2}T. \quad (7.15)$$

The minimum attainable variance is given by the sum of the two variances:

$$\text{var}(S) = \frac{S+B}{\frac{1}{2}T} + \frac{B}{\frac{1}{2}T},$$

and the relative standard deviation, σ_S/S , of the retrieved signal rate is

$$\begin{aligned} \frac{\sigma_S}{S} &= \frac{1}{S} \left(\frac{(S+B)}{\frac{1}{2}T} + \frac{B}{\frac{1}{2}T} \right)^{1/2} \\ &= (2/T)^{1/2} \frac{(S+2B)^{1/2}}{S} \end{aligned} \quad (7.16)$$

$$\approx \frac{2B^{1/2}}{T^{1/2}S}. \quad (7.17)$$

Note that, in arriving at (7.17), B enters twice in the derivation: once during signal measurement and once when background is counted. It follows from (7.17) that, to reduce the dispersion σ_S/S , it is theoretically more effective to increase S than to reduce B , since the latter appears as the square root in (7.17). For example, it may be more beneficial to enlarge an apparatus in preference to reducing the background, but there could be practical reasons, such as cost, for why this may not be feasible.

For example, suppose the signal rate is $S = 2$ per second, and the background is $B = 10$ per second. If we assume, on the basis of (7.15), that 50 s are allocated to each measurement, which is made just once, the relative standard deviations from (7.17) are

$$\sigma_S/S = \frac{2 \times 10^{1/2}}{100^{1/2} \times 2} = 0.316,$$

$$\sigma_B/B = (500)^{1/2}/500 = 0.045.$$

The results are therefore quoted as $S = 2 \pm 0.316$ per second, and $B = 10 \pm 0.045$ per second. Clearly, if higher precision is required, then, according to (7.17), to achieve a factor of 2 reduction in the uncertainty in S and B requires an increase of four times in T .

Minimal variance is achieved by allocating equal time to the measurement of $(S + B)$ and to that of B when performing photon counting on weak signals. This theory does not apply to the use of a DMM for making anode current measurements because the sample time is fixed within the instrument. However, a 50% duty cycle is achieved with more advanced DC instrumentation such as signal averages and lock-in amplifiers. Although superseded by other techniques, liquid scintillation counters were prominent in the past for measurement of tritium—a low-energy isotope of hydrogen. Anode outputs from two PMTs, operated in coincidence, are summed, producing a count rate expressed as the efficiency, E . A pair of quality PMTs has efficiency in excess of 60 % for the detection of tritium decays. Instrument performance is stated in terms of E^2/B , where E is the count rate derived from a standard vial of doped scintillator. This figure of merit is compatible with the $(S/N)^2$ formulation in (7.17) that shows S^2/B dependence.

7.5 Signal recovery instrumentation

7.5.1 Boxcar integrators

Signal recovery concerns the extraction of weak but repetitive signal from a sea of noise and background; the signal may be so weak that it is indiscernible from noise and background when viewed on an oscilloscope, for example. The boxcar averager charges a capacitor in a low-pass filter with a thin time slice of the input signal. Timing of this operation is controlled by a trigger pulse that defines the time origin, and initiates the generation of a delayed gate signal, t_g , of fixed width, Δt . The low-pass filter averages the signal after the set delay and for the set duration. Scanning t_g from zero to the full duration of the signal, many times, leads to visible recovery of the signal. The concepts involved are illustrated in Fig. 7.5, which represents the decay of a scintillator, or laser-excited fluorescence decay, for example. Freedom

7

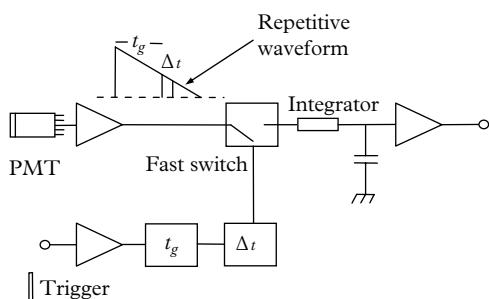


Fig. 7.5. The basics of a boxcar averager used to recover time structure in weak but repetitive signals.

from jitter in the trigger pulse is clearly important. The gated integrator amplifies and integrates whatever signal is present during the sample time, Δt . Any signal that appears while the gate is open adds linearly to that already accumulated in the integrator, while the noise adds as the square root of the number of shots.

7.5.2 Lock-in detection

The arrangement in Fig. 7.5 uncovers the time signature of repetitive signals, but a different configuration is required for the detection of low-level, random, light signals. Mixed in with the desired signal is an unwanted component, consisting of stray light, leakage currents, dark current, and amplifier noise. When the light source is blocked, only the unwanted sources remain. By alternatively blocking and unblocking the beam with a chopper, described below, we can generate a series of background counts b , interleaved with signal-plus-background counts, $s + b$. When the input to the PMT is cut off, the measured contributions include only the noise and background sources that lie between the chopper and the PMT. When the beam is unblocked, the background contributions remain in addition to the light signal. The essence of the lock-in method is the isolation of the signal through subtraction of adjacent samples. Subtraction is done electronically, while the division of the time base into equal time segments is performed by passing the signal beam through a rotating disc consisting of a multiplicity of symmetrical opaque vanes. As the disc rotates, it alternatively obscures and passes the beam, producing an on-off signal for detection by the PMT. Rotational speeds of typical commercial products are adjustable up to about 3 kHz in frequency, with a duty cycle of 50%. The chopping process converts a DC signal to an AC one with a fundamental frequency dictated by the angular velocity of the chopper. The signal recovery process is thus unaffected by any low-frequency contributions from 1/f noise (often referred to as drift) or by a slow change in background level (e.g. increase in dark current). Essentially, b is measured sufficiently close to the previous sample of $(s + b)$ that it is essentially a measurement of the background taken at the same time as the signal.

Low-light-level measurements with a PMT may be configured to produce an output in either of two possible forms: a sequence of distinct charge pulses, adequately separated in time, or as a DC level with superimposed amplitude fluctuations produced by noise. Electronic instrumentation is readily available to cover both these options. Although random, the average light flux is assumed constant over the duration of a measurement sequence, usually involving many samples. If the intensity of the light source varies over the duration of the entire experiment, which may extend to hours, then the measured result for S represents the average light flux over that period. An early synchronous system is that due to Arecchi *et al.* (1966) based on photon counting. Its elegance lies in conceptual simplicity: arguably, this is the most precise and accurate method available for making low-light-level measurements. Furthermore, the detector is easily realized by cascading a set of electronic modules (shown in Fig. 7.6), all of which are commercially available or easily constructed in an electronics laboratory.

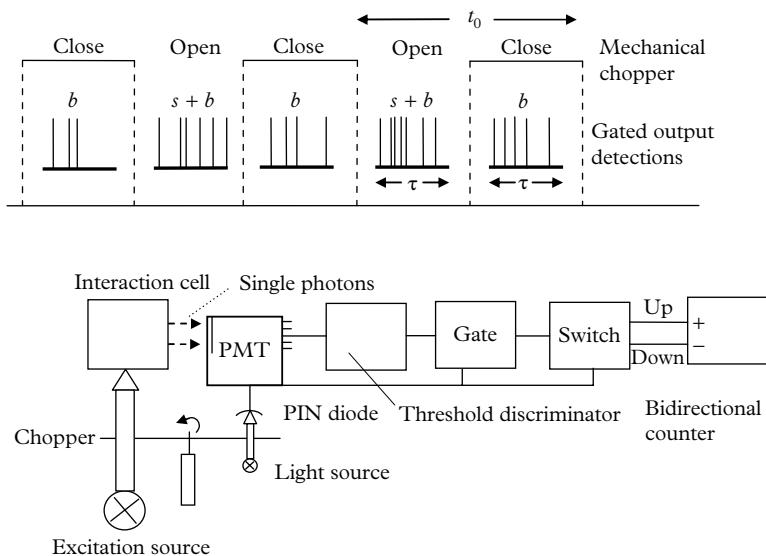


Fig. 7.6. The synchronous single-photon-counting detector based on the arrangement adopted by Arecchi *et al.* (1966). The upper diagram illustrates the action of the light chopper in alternatively exposing the PMT to signal plus background ($s + b$) and background alone (b). The bidirectional counter counts ‘up’ for the $s + b$ sequences and ‘down’ for b .

There is an analogue equivalent to the photon-counting instrument just described; in this, the anode current, or a voltage proportional to the anode current, is sampled by chopping. Signal updating and background subtraction is performed using a high-sensitivity voltmeter or electrometer. The analogue instrument benefits from AC coupling and thus enjoys the same degree of freedom from $1/f$ noise as previously mentioned. Statistical arguments always favour photon counting because the ENF does not apply. However, an analogue system modified to work on standardized digital signals from the output of an amplifier/discriminator would in theory offer the same resolution as straight photon counting. This is the mode of operation called ‘standardized storage’ by Jones *et al.* (1971).

The use of these sophisticated instruments has waned since the 1970s, primarily because of the advantages of fast, continuous digitizing techniques. Digital signal processing (DSP), based on continuous sampling with a fast analogue-to-digital converter (ADC), not only retains the original signal waveforms, but replaces analogue filtering operations with digital ones. There are primarily only two remaining manufacturers of signal recovery instrumentation: Stanford Research Systems (SRS), and Advanced Measurement Technology (Ametec), the latter company having subsumed the products of other past, well-known manufacturers, such as Ortec Brookdeal, EG&G, and Perkin Elmer.

7.5.3 Synchronous signal averaging

The detection methods about to be considered are particularly suited to the extraction of a weak signal buried in a much higher background count. As an illustration, we consider the situation in which the background is of the order of ten times the signal. Following (7.15), counting times should be equal. Two reported examples of signal recovery are of particular merit: Arecchi *et al.* (1966) describe the extraction of $S = 1.1$ counts/s from $B = 33.5$ counts/s, with a precision in S of 28%; Oliver and Pike (1968) recovered $S = 0.043$ counts/s from $B = 0.459$ counts/s, with a precision of about 33%. We know from 7.5.10 that the counting statistics will be unfavourable to the achievement of high precision, unless the total measurement time, T , is long. Arecchi *et al.* achieved the stated precision after 24 minutes; Oliver and Pike's measurement took 2 hours to achieve a similar performance.

The observation concerning the benefits of interleaving signal and background measurements leads to the appropriately named 'synchronous photon detection method'. Consider a detector alternatively exposed to a light source and to background by chopping. This sequence is repeated many times, as depicted in Fig. 7.7. Following the treatment outlined by Oliver (1977), background counts are recorded for the first period as n_1 , the signal-plus-background counts as n_2 , the next dark counts as n_3 , the following signal plus dark counts as n_4 , and so on. Successive samples of signal counts are obtained from the difference estimators,

$$\begin{aligned}s_1 &= (n_2 - n_1), \\ s_2 &= (n_4 - n_3), \\ s_3 &= (n_6 - n_5), \\ &\dots,\end{aligned}$$

and similarly for background:

$$\begin{aligned}b_1 &= (n_1 - n_3), \\ b_2 &= (n_5 - n_7), \\ b_3 &= (n_9 - n_{11}), \\ &\dots\end{aligned}$$

The signal estimators sum to the total count for a period of $T/2$, while the background estimators sum asymptotically to zero. This method of data handling

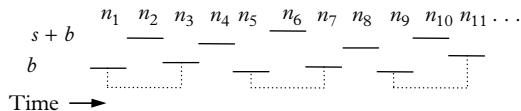


Fig. 7.7. The sequence of measurements n_1, n_2, \dots, n_k . Background estimators b_1, b_2, \dots, b_k follow from the readings linked by dotted lines. Note, in this particular arrangement, there are twice as many signal indicators as background ones.

provides $n_k/2$ estimators of the signal, and $n_k/4$ for background. Means and variances for these estimators follow from (7.16):

$$\begin{aligned}\langle s_k \rangle &= \frac{1}{2}ST, \\ \langle b_k \rangle &= 0, \\ \text{var}(s_k) &= \frac{1}{2}ST + BT, \\ \text{var}(b_k) &= \frac{1}{2}BT\end{aligned}$$

A synchronous single-photon-counting configuration, chosen by Arecchi *et al.* (1966), is given in Fig. 7.6. The alternate exposure and blocking of the light is performed by a light chopper, of frequency $1/t_0$, comprising a symmetrical pattern of opaque, wedge-shaped segments in a windmill pattern. The speed of rotation is selectable but for long time cycles, lasting from a few seconds to minutes, one may substitute a stepper motor drive for the continuous one. The sharpness of the transition between signal-on and signal-off depends upon the rotational speed, the width of the beam, and on any jitter in the mechanics. Arecchi *et al.* were able to eliminate this source of timing uncertainty by opening an electronic gate for a fixed time, $\tau < \frac{1}{2}t_0$, nested wholly within the chopping window. The signal for this operation derives from the fixed light source and photodiode combination shown in the figure. This particular arrangement directs the counts alternatively to the ‘add’ and ‘subtract’ inputs of a bidirectional counter, yielding the net signal counts only. Information on the behaviour of the background over the measurement period is therefore lost, and establishing the mean level and stability of the background requires a separate run with the light source disabled. This is not entirely satisfactory for the reasons mentioned earlier: the background measurement should ideally be made immediately before or immediately following the measuring signal.

7.5.4 Signal recovery simulation

Signal recovery can be simulated by generating random numbers with a Poisson distribution, for both signal and background, subject to assumed mean values for S and B . The exercise readily provides an indication of the time required to separate a feeble signal from a much larger background count. This is the best case, but it is useful when planning an experiment, although in practice the statistical fluctuations will exceed theoretical predictions. Successive counts for $s + b$ and b are assigned to the sequence n_k , enabling the extraction of the total signal counts with the passage of time. The evolution of the two estimators is shown in Fig. 7.8 as they are progressively updated. Data points perform a random walk about their mean values, constrained by the standard deviation parabolas representing the noise. Note that curves for one standard deviation are indicated but this does not imply that all the data points will remain within these boundaries. The random-walk nature of the data time lines means that every simulation will be different for repetitive measurements. It takes about 70 min

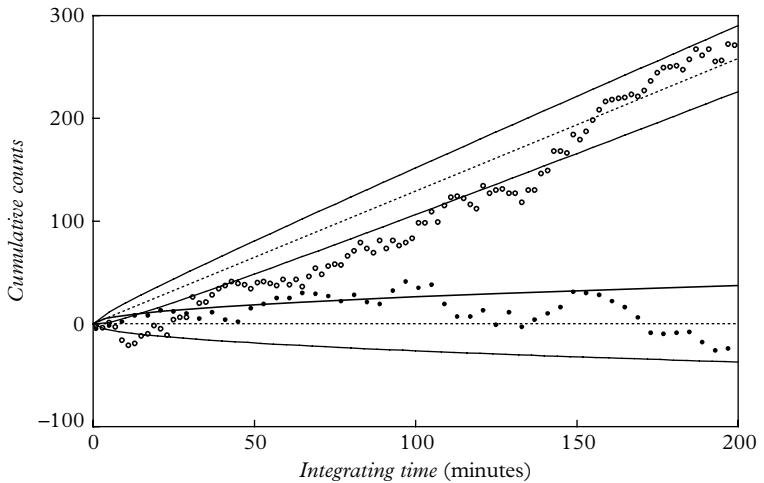


Fig. 7.8. Simulated recovery over time of a weak signal (open circles) of 0.041 counts/s from a background of 0.459 counts/s (solid dots). Actual plots of this form are given by Oliver and Pike (1968) and Arecchi *et al.* (1966).

7

for the signal path to separate from the background noise for the simulation shown. The obvious advantage of subdividing the time in this way is gained by using the local background for subtraction purposes. To facilitate comparison of experiment with theory, the background and signal rates for Fig. 7.8 were set identical to the experimental values used by Oliver and Pike (1968).

7.5.5 Subdividing time intervals

In theory, calculated values for mean and variance are independent of the division of T into multiple sub-intervals: the answers always turn out the same. For example, the variance of 100 samples of 1 s each is the same as that for a single exposure of 100 s (this follows from (4.34)). However, it is bad practice to rely on a single measurement of $s + b$ and a single measurement of b . There are multiple reasons why taking a measurement over a long time duration may lead to systematic uncertainty: first, the background may vary over the course of measurement because of temperature changes or drift in dark counts with time; second, the source strength may vary with time; third, the sensitivity to interference, the bane of low-level-signal experiments, may be masked by long exposures. Expanding on this last point, electromagnetic interference from mains switching and electric drills and other equipment can be pinpointed in time by taking readings at sufficiently close intervals, otherwise missed for long exposures. These sources may not be so evident, if the time bins are long. Undoubtedly, in this type of investigation, measurements would be acquired by a data logger, allowing the experimentalist a wide choice of appropriate subintervals. The practice of

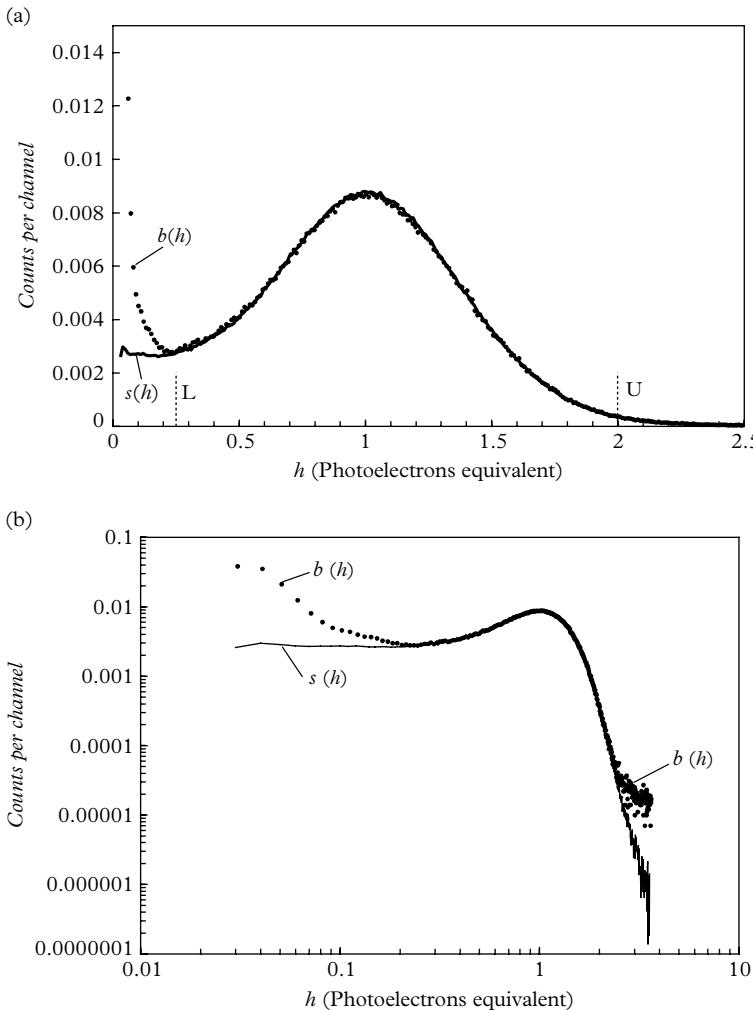


Fig. 7.9. (a) Normalized, differential pulse height distributions for single-photon excitation, $s(h)$, and for the dark count, $b(h)$. Following convention, the channel in which $s(h)$ attains a maximum is taken as one photoelectron equivalent. (b) Log-log plots of the data in the top panel uncovers multi-photoelectron background pulses obscured in the linear presentation.

7

eliminating high-count readings on the assumption that they are spurious must be exercised with caution, making allowance for the fact that small numbers are subject to high fluctuation. Only those events that are considered genuinely affected by interference should be eliminated. If the mean count per unit time interval is 10, say, then a count of 500 is certainly spurious; on the other hand, one

of 20 counts may be genuine. This kind of data manipulation is of doubtful validity, and the best solution lies in removing the problem at the source, rather than attempting to deal with its outcome.

7.5.6 Upper and lower thresholds

The functional dependence of $b(h)$ has an important bearing on setting the lower and upper levels, L and U. As previously mentioned, the profile of the background pulse height distribution, $b(h)$, is dissimilar to that of signal. The essential differences can be stated as follows, for normalized pulse height distributions: $b(h)$ has a higher content of small pulses, and multi-photoelectron pulses, compared with $s(h)$, as illustrated in Fig. 7.9. The sources of sub-photoelectron and multi-photoelectron pulses are discussed in 6.6 and illustrated in the differential distributions of Figs 6.1 and 6.2.

Settings for L and U should be compatible with the SER of the chosen PMT. Normalized integral pulse height distributions, shown in Fig. 7.10 for signal and background, provide a guide to realizing a suitable setting. Between 80 and 90% of the counts are included, following the suggested setting of L in the region of 0.25 photoelectron equivalent. Note that the valley in the differential distributions $s(h)$ and $b(h)$ occurs at L in this selection. The choice of L coincides with the most stable point on the differential distribution. On the high pulse height side, for both signal and background distributions, we see that Region C (see Fig. 6.1) contributes at the 1 % level by setting $U = 2$. This procedure for setting threshold levels highlights the difficulty in attempting absolute light measurements—that is,

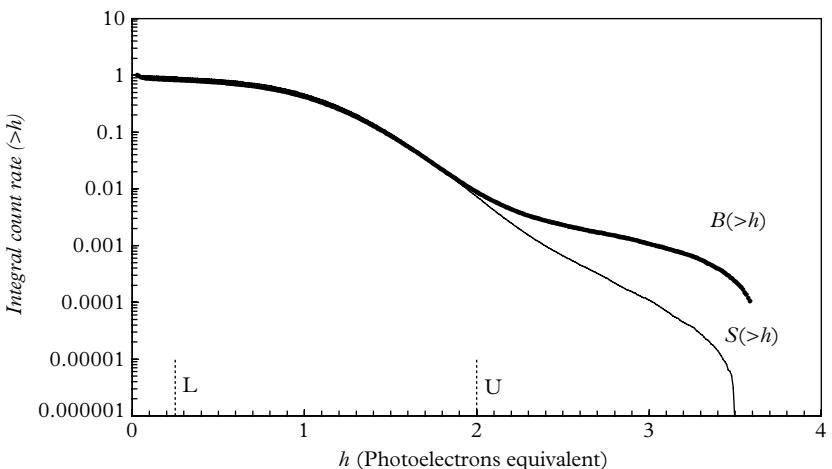


Fig. 7.10. Normalized, integral pulse height distributions, for single-photon excitation of the photocathode, derived from Fig. 7.9 (b). Suggested upper, U, and lower, L, discriminator settings at $\frac{1}{4}$ pe and 2 pe, respectively, are also shown.

relating count rate to photon flux—since the choice of thresholds depends on user judgement and predilection. The PMT chosen in this example is one of quality but the procedure is more uncertain for a PMT with a poorer SER.

7.5.7 Signal magnitudes: Photon counting

It is important in some applications to be able to predict the expected magnitude of signals produced by single-photoelectron excitation. For the moment, it is assumed that each single photoelectron can be amplified sufficiently for conventional electronics to detect; this will indicate amplification by at least a factor of 10^6 . An equivalent output circuit of a PMT is a current generator feeding a parallel combination of a resistor and a capacitor. There is a limiting capacitance, referred to as C_o ; it comprises capacitance between the anode and other elements of the PMT, together with a contribution from the socket; C_o is typically 10 to 20 pF and its presence can have an unwelcome broadening effect on the shape of fast pulses produced at the anode, and hence on the count rate capability. One way of overcoming the effect of C_o is to choose R small, for example 50Ω , in which case the RC time constant is also small and of the order of 0.5 ns—this is much smaller than the t_{fwhm} of any conventional PMT. Consequently, the peak amplitude of the voltage pulse will be determined primarily by the 50Ω resistor, in accordance with (7.18) and Fig. 7.11. Note that this value of R is compatible with the characteristic impedance of standard coaxial cable for signal transmission—RG174, for example.

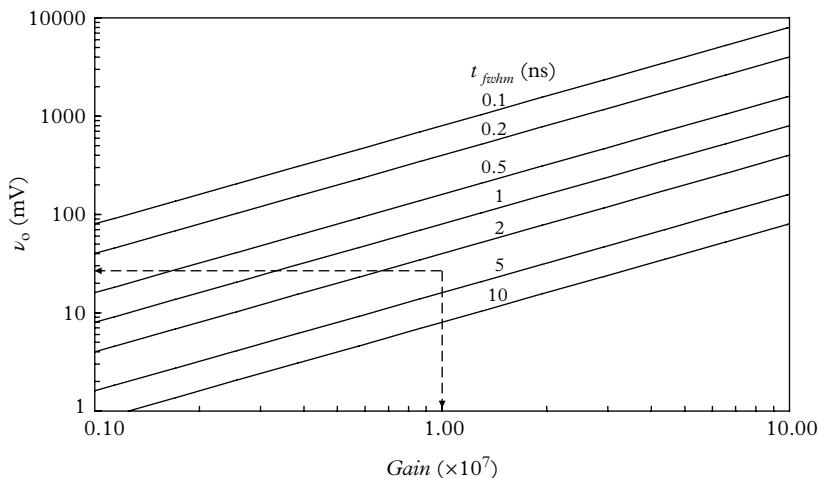


Fig. 7.11. The relationship between PMT gain, speed of response, and peak output signal, ν_o , from (7.19) for single-photon excitation with $R = 50 \Omega$. The dashed lines define the performance for a typical fast PMT with $t_{fwhm} = 3$ ns.

The gain required of a PMT used for photon counting must be sufficient to match the sensitivity of available electronics—either commercial or that devised by the user. It is shown in 7.6.1 that high-gain photon-counting-quality PMTs tend to perform optimally within the gain range of 10^6 to 10^7 . Does this gain generate a signal of sufficient magnitude for detection on an oscilloscope, or some other instrumentation of low cost? This is easily tested by making use of (7.18), which predicts a signal of 27 mV for $\langle g \rangle = 10^7$, where $t_{fwhm} = 3$ ns, and with $R = 50 \Omega$, for example. This signal level is easily detected by any reasonable oscilloscope but is barely sufficient for a NIM series discriminator of 50 mV threshold. It is therefore common practice to include a fast amplifier with a gain of 10 to set the desired threshold within the millivolt range:

$$v_0 \approx 1.6 \times 10^{-19} \langle g \rangle R / t_{fwhm}. \quad (7.18)$$

The family of curves in Fig. 7.11 show the obvious disadvantage of using slow Venetian blind or box-and-grid PMTs for photon counting: the peak amplitude of signals is reduced to a few millivolts, which implies either the need for more sensitive electronics or an increase in HV. Amplifier discriminators with a standard threshold of 2 mV are available from several manufacturers, and some units are adjustable over the range of 1 to 5 mV. Attempting to set a discrimination level below 1 mV when using low-gain PMTs invariably leads to unstable performance, covered in greater detail in 7.6.5. Extending the discussion to other tube types, we see that the well-established box-and-grid and Venetian blind PMTs, with their relatively slow time responses, will require operation at a gain of $\sim 4 \times 10^7$. The necessary gain for an MCPPMT will be less than 10^6 because of its comparatively fast time response; all other PMT types fall within a gain range of 10^6 to 10^7 for photon-counting purposes.

7

7.5.8 Signal magnitudes: DC detection

An approach similar to the one in 7.5.7 may be followed to establish the required sensitivity for DC detection of a low-level signal. An input of n pe/s will produce an output voltage of

$$v_0 = en\langle g \rangle R, \quad (7.19)$$

where the symbols have their previous meanings. Equation (7.19) is represented in Fig. 7.12, with R as a parameter. Consider, for example, operation at low light levels by assuming $n = 100/\text{s}$, $\langle g \rangle = 10^7$, and $R = 1 \text{ M}\Omega$; (7.19) predicts a DC level of only 0.16 mV, which immediately highlights the difficulties in low-light-level measurement based on this method. Capacitance may be added to smooth the output but without affecting the mean voltage. Measurements made with a multimeter or oscilloscope would need signal levels of the order of 100 times this level. Increasing the PMT gain to 10^9 is not an option nor is increasing the load resistor and hence the detector time constant.

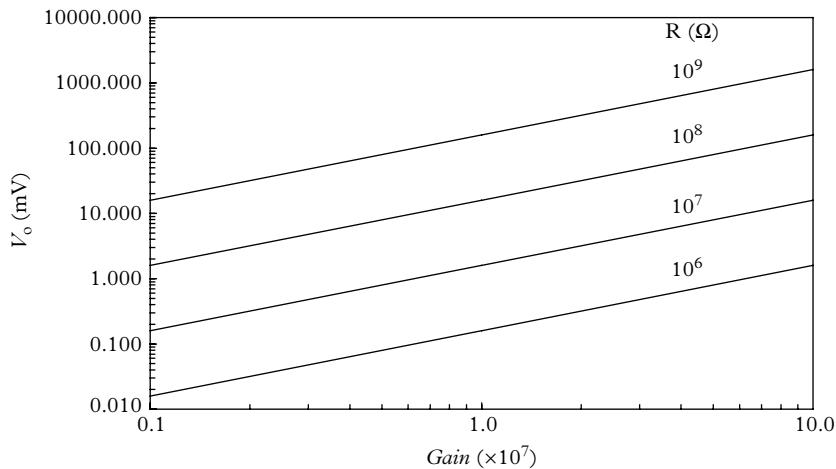


Fig. 7.12. Current-to-voltage conversion from anode current to DC voltage for a fixed cathode emission of 100 electrons/s. The resistor may be a standalone anode load, or possibly integrated into the input electronics (see also Figs 7.23 and 7.24). The high-value resistors required to generate millivolts should be noted.

7.5.9 Excess noise factor

Two essential pieces of equipment are required to perform this calibration efficiently: a source of single photons and an MCA. In theory, an SCA or a threshold discriminator could be used for this purpose, but gathering the necessary data becomes tedious. A single-photoelectron source is easily devised, whether random or synchronized, by following the method discussed in 4.1.1. Wherever possible, the characteristics and disposition of the calibration light source should mimic the actual application, with respect to:

- the wavelength of light, and its polarization
- the area of illumination on the PMT window
- geometrical effects, such as the angle of incidence
- possible coincidence requirements
- operation of the PMT at the same gain as in the application

This is because the shape and hence the associated variance of the SER distribution is sensitive to all the foregoing. Certain commercially available MCAs include a gating facility, which provides the option of recording a spectrum subject to an external coincidence condition—such as the trigger pulse from a laser. If synchronous detection is employed, the SER should be recorded under such conditions

because the variance tends to be lower compared with that obtained during asynchronous operation.

The distribution recorded by an MCA is essentially a record of the counts in each channel as a function of the channel number—of which there are typically 1024. Each channel number, Ch , may be expressed in terms of a charge $Q(Ch)$, through calibration of the instrument. However, ENF is dimensionless, and charge calibration is not mandatory for its calculation. It is important that any offset in the MCA should either be removed or the abscissa of the distribution corrected. The mean and the variance are calculated in the standard way and, for a well-resolved SER, the mean of the distribution will lie close to the most probable; the ENF ranges from close to unity for SiPMs and HPDs to about 1.5 for Venetian blind PMTs. Typically, PMTs used in high-energy physics and for industrial photon-counting-based instrumentation will have a noise figure of between 1.1 and 1.2. Note, the ENF for a background spectrum, $b(h)$, is determined in the same way as for signal.

To summarize, pulse counting offers the possibility of enhancing the signal-to-background ratio through eliminating background pulses located in the extremities of the SER. This brings the additional benefit of reduced variance, because a significant number of multi-photoelectron pulses are correlated. In photon counting, we are able to eliminate the contribution to variance from the noise factor term in whichever formulations it appears. Every analogue pulse falling within the discriminator window is given unity weight, and counting these pulses achieves the ultimate in low-noise detection: the variance in the count rate is simply $1/(\eta n T)$ —the lowest limit possible by any detection method. Strictly, this limiting variance is not attainable because of the contribution from threshold jitter in the discriminator—another source of noise. This effect is illustrated in Chapter 8, Fig. 8.25.

7

7.5.10 Ratemeter operation

Photon counting can be described as the generation of standard pulses (e.g. transistor-transistor logic (TTL)), with digital storage of the number; anode current measurement (DC) generates non-standard pulses for storage on a capacitor. A rate meter combines the two techniques by replacing non-standard pulses by standard ones. Noise associated with charge storage on the capacitor is now a function of detection rate only. Standardization is achieved by interposing a threshold discriminator between the PMT and the RC circuit. A discriminator-monostable combination generates the required standard pulse for each PMT signal that exceeds a set threshold. This seems a major advance since we can overcome the influence of the noise factor in DC measurements by reverting to a detection scheme similar in principle to a photon-counting one.

Consider an ideal PMT, with a noiseless multiplier, observing a constant flux of photons producing n pe/s, on average. Amplified charge, Q , is collected by a capacitor, C , that is shunted by a resistor, R . The voltage on the capacitor, when viewing a steady light source, will reach quasi-equilibrium after the elapse of about four time constants. If the light intensity is changed, a new steady state will be

attained, again after about four time constants have elapsed. These equilibrium states have defined but noisy count rates, as depicted in Fig. 7.2(c). The derivation of (7.20), the expression for variance, was originally devised for nuclear radiation meters but applies equally to single-photon detection. The derivation will not be repeated here because it is adequately covered in the scientific literature (see e.g. Knoll 1989 and Jones *et al.* 1971). Equation (7.20) refers to the standardized mode through the use of a discriminator:

$$\frac{\text{var}(\bar{Q})}{\langle \bar{Q}^2 \rangle} = \frac{1}{2\eta n\tau} = \frac{2}{\eta \mu T}. \quad (7.20)$$

In this final step, $\tau = RC$ has been taken equal to $T/4$, where T is the sample time; selecting a time constant of τ equal to four sampling times is somewhat arbitrary but practical. The ‘non-standardized’ mode of operation in Fig. 7.2(c) includes the effect of noisy gain, described by (7.21):

$$\frac{\text{var}(\bar{Q})}{\langle \bar{Q}^2 \rangle} = \frac{2}{\eta n T} \times \left(1 + \text{var}(g)/\langle g^2 \rangle \right). \quad (7.21)$$

Recall from 4.78 for photon counting that output pulses are standardized, and

$$\frac{\text{var}(\bar{Q})}{\langle \bar{Q}^2 \rangle} = \frac{1}{\eta n T}. \quad (7.22)$$

Excess noise factors range from 1.2 to 1.5 and, when included in (7.21), increase the relative variance by the factors indicated. A system based on capacitive storage would therefore require more than twice the time needed with digital storage to achieve the same noise performance.

A constant flux causes the capacitor to reach quasi-equilibrium after about four time constants, with what is referred to as a 2% exponential residual. This is because, after four time constants from switch-off or switch-on, the signal is still at 2%, $\exp(-4)$, of its previous level; clearly, the system has memory at the 2% level, which implies correlations. The limitations of this method for measuring light levels are twofold: first, contiguous samples are correlated and, second, the degree to which a changing light level can be followed depends upon the time constant, τ , chosen. Note in both (7.20) and (7.21) we can reduce the relative variance by extending T . A worked example examining the performance of a laboratory ratemeter, comprising a DMM shunted by an RC network, helps in the illustration of these formulae:

photon rate: $n = 4 \times 10^6$ photons per second

QE: $\eta = 25\%$

multiplier gain: $\langle g \rangle = 10^6$

anode load: $R = 1 \text{ M}\Omega$, $C = 1.0 \mu\text{F}$

$T = 4\tau = 4$ seconds

anode current: $I_a = 1.6 \times 10^{-19} \times 4 \times 10^6 \times 0.25 \times 10^6 = 1.6 \times 10^{-7} \text{ A}$

$$\text{anode voltage: } \langle V_a \rangle = 160 \text{ mV}$$

$$\text{assumed ENF} = \left(1 + \frac{\text{var}(g)}{g^2}\right) = 1.2$$

$$\text{var}(V_a)/\langle V_a \rangle^2 = 2 \times 1.2/(4 \times 10^6 \times 0.25 \times 4) = 0.6 \times 10^{-6} \text{ using (7.21)}$$

$$\sigma(V_a)/\langle V_a \rangle = 0.08 \% = \pm 0.13 \text{ mV}$$

The fluctuation of 0.13 mV would manifest itself as a flickering of the least significant bit of a 3½-digit instrument, sampling every 4 s. If we were to count the number of pulses n over 4 s, the noise factor is unity and we would have

$$\sigma(n)/n = 1/n^{1/2} = 1/(4 \times 10^6)^{1/2} = 0.05 \text{ %}.$$

A point worth making here, although it applies more to nuclear radiation monitoring than to photon counting, concerns the interpretation of the output signal. For a standardized instrument, the output reading is proportional to the number of detections per unit time, regardless of their energy. In non-standardized operation, the energy deposited per second is measured and, in this case, the instrument performs analogously to a dose meter.

7

7.6 PMT selection criteria

Noise considerations, covered in 7.5.10, indicate that choosing a PMT with a low ENF for both signal and background, contributes to optimal signal-to-noise performance. For a particular PMT, the capability to detect low light levels obscured by dark count noise and by the noise in signal depends on the signal-to-noise ratio for that particular PMT. An additional consideration is the signal-to-background ratio, the maximization of which depends on several parameters. These two ‘quality of performance’ ratios must never be taken as synonymous—they refer to different aspects of noise.

Following the treatment in 7.4.1, the first priority in both analogue and photon counting applications should be to increase the signal S by:

- (1) concentrating the light from the source on to the photocathode
- (2) selecting a high-QE photocathode, ideally spectrally matched to the light source

and to reduce the background B by:

- (1) choosing a PMT with cathode dimensions to match the area of illumination
(Changing tube type to one of smaller cathode diameter generally leads to a fall in dark count. The actual reduction factor lies somewhere between the ratio of the diameters and the ratio of the areas. PMTs with reduced-area

photocathodes (active diameter 2.5 mm), such as the 9863/100 and 9893/100 types, also warrant consideration.)

- (2) paying a premium for a selected PMT
- (3) using a combination of selection and cooling, which could lead to an improvement by a factor of 5 for bialkali PMTs, while the improvement factors for multi-alkali and other infrared-sensitive detectors are considerably higher
- (4) reviewing the choice of photocathode type

The next concern is that of choosing the optimum operating voltage V , for which there are three considerations. Manufacturers provide a recommended operating gain, which they call the nominal gain, for all PMTs. They also recommend a maximum gain that is usually ten times the nominal. Ideally, one should operate within a factor of 10 on either side of the nominal gain. If it transpires that operation is outside these limits, it implies that the original choice may not be the right one and should be reviewed. Second, the sensitivity and dynamic range of the measuring electronics must be considered in terms of the maximum current capability of the PMT. Manufacturers stipulate $I_a(\text{max})$, with limits generally lying within the range of 10 to 100 μA , depending primarily on the area of the dynodes, and therefore indirectly on tube diameter. There are two undesirable effects related to drawing high anode currents: if the anode current is excessive and continuous, the PMT will rapidly show signs of fatigue, and the gain will generally decrease with time. Admittedly, the gain can be restored by adjusting the voltage, but this is not always desirable or convenient. At high and variable anode currents, there are also practical difficulties in maintaining fixed bias voltages on the dynodes with an all-resistor divider—failure to do this leads to gain shift.

7.6.1 Optimal gain for DC detection

The quality of detection attainable in low-light-level applications depends on the method adopted. A comparison of three basic detector configurations is considered in 7.4: photon counting, DC detection, and shot noise power measurement. The exercise considers how to optimize detection in terms of signal-to-noise ratios. It is essentially a statistical consideration in low-level-light detection, but maximizing signal-to-background ratios may be even more important.

Cooling is an obvious way to reduce background, especially if the photocathode is red sensitive. But this is not always feasible because of cost and space considerations. In an ideal PMT, signal and the dark current increase linearly with gain, in which case there is no optimal setting of the voltage regarding signal-to-background ratios. The source of dark current is such that it varies non-linearly with gain, over a significant range of HV. This is because it comprises a leakage component, I_L , and an amplified photocathode dark current, $\langle g \rangle I_k$. The combined anode dark current is $I_B = I_L + \langle g \rangle I_k$, where we note that leakage current is a function

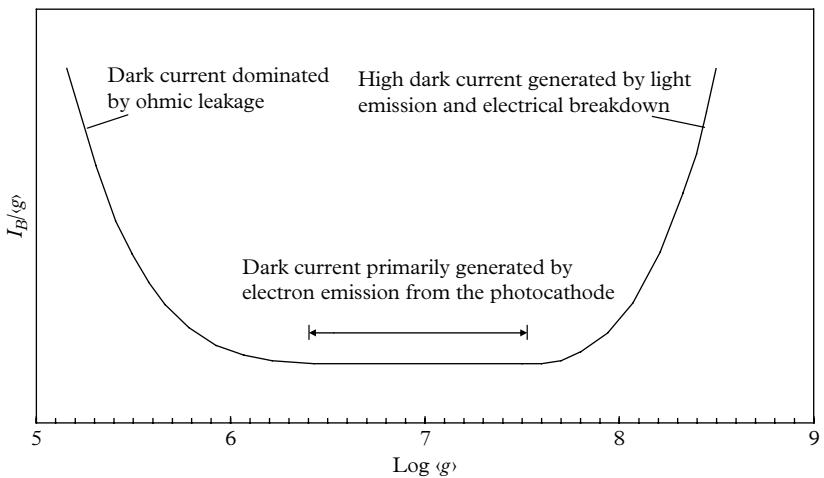


Fig. 7.13. A graphical means for locating the range in gain over which the signal-to-background ratio is maximized. Data should be taken at the intended temperature of operation.

7

of HV whereas the pulsed component, $\langle g \rangle I_k$, varies linearly with gain and hence as a power of HV.

At low voltages, and therefore at low gain, the leakage component generally exceeds the pulsed one, and the total dark current varies roughly in proportion to the HV. At sufficiently high gain, the reverse holds and the total dark current varies approximately linearly with gain. This quasi-linear dependence of I_B on gain continues until the onset of electrical breakdown, as a result of which an erratic and non-linear increase in I_B is observed. Three distinct regions of PMT behaviour, illustrated in Fig. 7.13 are clearly delineated in an actual PMT, in Fig. 7.14, which is taken from Fig. 6.6. The resulting curve indicates the gain range over which best performance is achieved, and also serves as a measure of PMT quality: the lower the minimum, in terms of equivalent cathode dark current, the better is the quality of the PMT. It is worth noting with regard to Fig. 7.14 that knowledge of absolute gain is unnecessary for setting-up purposes: it is sufficient to use relative gain, a parameter easily measured in any experimental arrangement.

7.6.2 Optimal HV for photon counting

A plateau characteristic curve is one of four interrelated distributions, linked to voltage, V , and pulse height, h . These functional relationships are explored in detail in Appendix A, and the reader may wish to refer to the mathematical treatment at this stage, before moving on to practical considerations. The integral distribution $S(>h_0, V)$ for a fixed threshold h_0 and variable V represents the contribution from all signal pulses in the spectrum that exceed the fixed threshold h_0 .

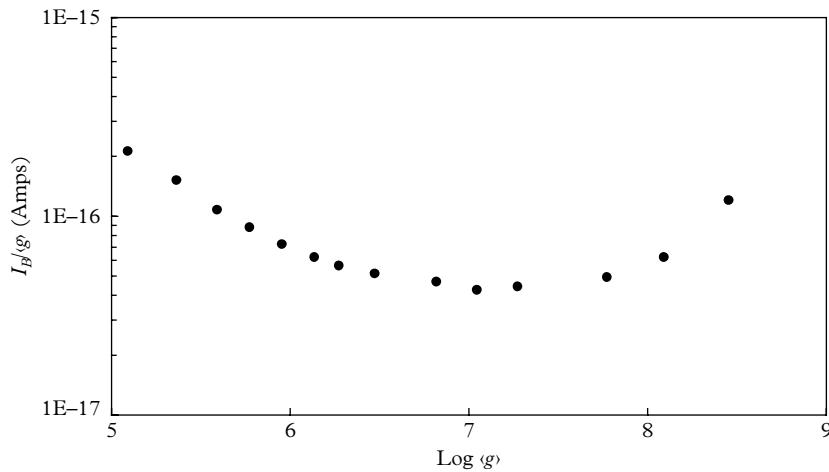


Fig. 7.14. The optimal operating gain lies in the vicinity of the minimum of the curve. Dividing I_B/g by the electronic charge gives the number of electrons per second: that is, an optimal dark count equivalent of 265 electrons/s at $g = 10^7$. An ideal PMT is one that is free from leakage currents and with dark counts independent of voltage; it would be represented by a horizontal line in the figure.

7

It is a cumulative count rate distribution and, if h_0 is set at zero, $S(>0, V)$ represents the integral count rate as a function of V . In practice, of course, h_0 has to be set sufficiently above zero to attain meaningful and stable performance. The precise location of the optimal setting cannot be stated unambiguously: this is the investigator's prerogative, but there are guidelines.

Setting the optimal operating conditions is a prerequisite for performing successful photon counting. All instrumentation techniques share a common requirement of setting the HV. Measurements based on counting anode pulses that exceed a fixed threshold are based on 'a plateau characteristic curve'. Detection by this method finds wide application, essentially because of its simplicity. An example of such a curve is given in Fig. 7.15, the features of which are explained as follows:

Region (a): Few pulses in the spectrum of pulse heights have sufficient amplitude to exceed the threshold, and count rates are correspondingly low.

Region (b): The count rate at 700 volts is approximately half the counts in Region (d) and therefore it corresponds to the median of the SER. Since we are dealing with single-photoelectron events, we can assign a scale centred on (b), on the basis that 1 photoelectron equivalent corresponds to the 'half plateau' count rate at (b). Noting that one photoelectron corresponds to 700 volts and that the gain for this particular PMT varies as $\sim V^{9.4}$ leads to the scale shown along the top of Fig. 7.15.

Region (c): The knee voltage, V_{knee} , is a further parameter of particular interest to manufacturers: it heralds the approach of the flat region of the characteristic and serves as a guide to selecting the operating voltage. V_{knee} relates to the relative slope, $(1/S) dS/dV$; based on experience, V_{knee} may be taken as the voltage at which the calculated slope attains $1.0\%/\text{Volt}$. An operating point at $V_{\text{knee}} + 100 \text{ V}$ corresponds to a relative slope of $\sim 0.1\%/\text{Volt}$ in this example. The method is clearly subjective, but the slope criteria may be adjusted to better suit a particular PMT. This procedure forms the basis for manufacturer's test purposes and user applications.

Region (d): The slope of the characteristic is reasonably constant at $\sim 0.1\%/\text{Volt}$ between 870 and 1200 V—termed the ‘plateau length’. Operating within Region (d) mitigates any loss of PMT gain, and the effect of divider instability with count rate.

Region (e): There is no justification for operating within this region because of the increasing slope and enhanced signal current. Note that the discriminator threshold approaches one-thousandth of a photoelectron at the extreme end of Region (e), clearly an unacceptable region of operation.

7

The photoelectron scale at the top of the diagram reveals the rapid decrease in threshold with increasing HV .

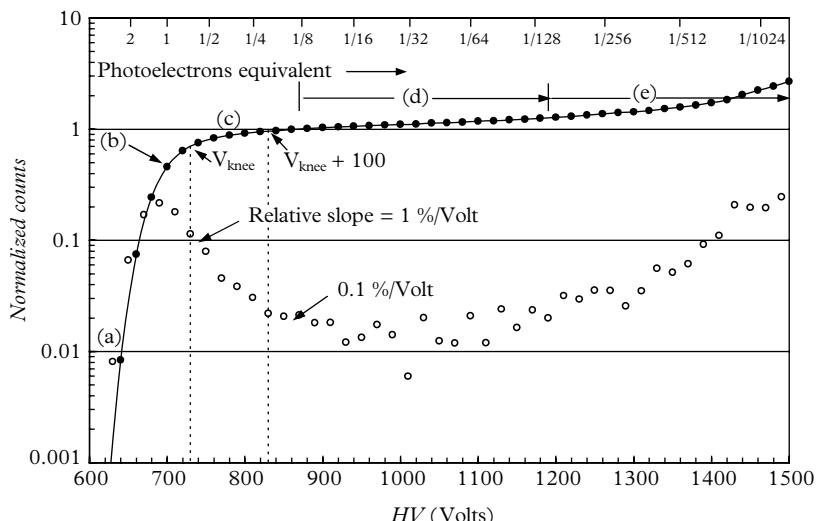


Fig. 7.15. A typical normalized plateau characteristic (●) for a 9107 photon-counting PMT showing features (a) to (e) referred to in the text. Each data point corresponds to the count rate of pulses that exceed a fixed threshold. The difference between adjacent readings, when plotted at their intermediate HV, is shown by circles. This highlights the observation that the plateau never attains zero slope and justifies the choice of operating voltage. The photoelectron scale is explained in the text.

7.6.3 MCA plateau characteristic

Photon counting may be performed with an MCA by utilizing both its differential and its integral modes of operation. This instrument offers a superior method of measurement because the differential spectrum, $s(h, V_0)dh$, is always on display. Being able to see detailed spectral information, from which the integral derives, is both a useful and a reassuring diagnostic. The starting point is a pulse height distribution for single-photon excitation of constant intensity. This is shown in Fig. 7.16 and labelled $V_0 = 860$ V. Pulses exceeding $h_0 = 25$ channels constitute the count rate $S(>h_0, 860)$. An immediate observation at low gain, and consequently low operating voltage, V_0 , is how few pulses have sufficient amplitude to exceed h_0 . As the voltage is increased to $V_1 = 935$ V, a higher proportion of the distribution—approaching 50 % in this case—is counted. Further increases in V result in higher counts but the rate of increase decreases until, at $V_4 = 1190$ V, virtually all single-photon initiated events are counted. The set of readings, $S(>h_0, V_0), S(>h_0, V_1), \dots, S(>h_0, V_n)$, taken over the range of fixed HVs V_0, V_1, \dots, V_n , constitute the plateau characteristic in Fig. 7.17:

$$S(>h_0, V_n) = \int_{h_0}^{\infty} s(h, V_n) dh, \quad \text{with } n = 0 \text{ to } n. \quad (7.23)$$

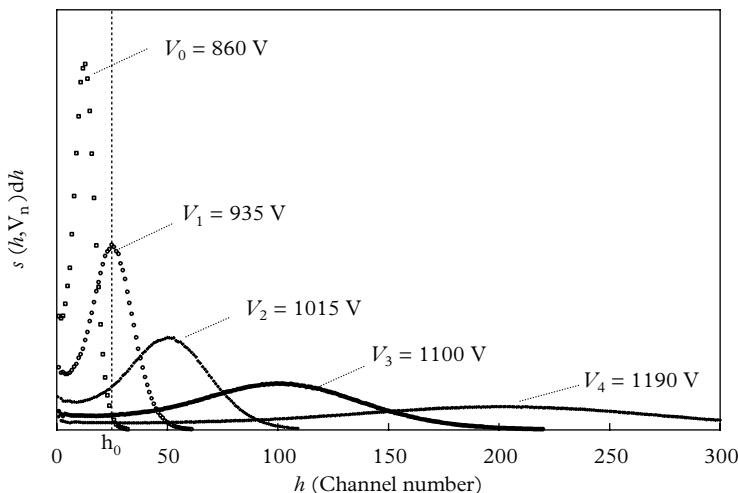


Fig. 7.16. A series of differential, single-photoelectron pulse height distributions, $s(h, V_n)dh$, measured with an MCA. Each distribution is taken at the indicated voltage. Note that, at the low-gain setting of 860 V, few signals exceed the threshold h_0 , which is set at Channel 25.

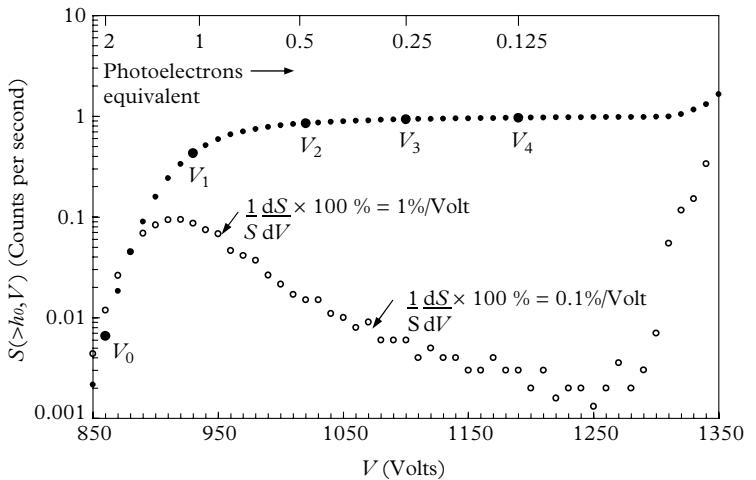


Fig. 7.17. Normalized plateau characteristic derived from Fig. 7.16: the bold dots refer to the five distributions $s(h_0, V_n)dh$, and the remaining points are taken from intermediate pulse height data in increments of 10 V. Circles represent increments in S , obtained by subtracting adjacent S readings. The two arrows point to positions on the incremental curve, ΔS , where count rate stability, $(1/S)(dS/dV)$, is $0.1\%/\text{V}$ at 1065 V, and $1\%/\text{V}$ at 960 V.

7

Note in (7.23) that the area under each differential distribution, with h ranging over 0 to ∞ , is common, because the light source is constant in intensity; V_0 to V_4 link the plateau counts to the corresponding pulse height differential distribution. Constructing the photoelectron scale at the top of the figure follows a similar method employed in Fig. 7.15, but using the peak positions of Fig. 7.16 for scaling. Although h_0 , stated in terms of pulse height or charge, is fixed, the threshold in equivalent photoelectrons decreases as V increases. For example, by operating at $V_1 = 935$ V, all pulses greater than 1 photoelectron equivalent (the peak of the SER) are counted. By operating at $V_2 = 1015$ V, all pulses of amplitude exceeding 0.5 photoelectron equivalent are counted, and so on. The plateau extends beyond V_4 , the highest voltage shown in the figure, but measurements are terminated on the approach of breakdown in the PMT, which is evident in Fig. 7.17 by the upturn in counts that starts at 1300 V. This rapid increase in counts is also evident in Fig. 7.9.

7.6.4 Critical PMT parameters

Following standard practice, plateau characteristics are invariably presented on a logarithmic ordinate scale, as in Fig. 7.18(a). The considerations in 7.6.3 indicate an operating point located at 890 V in Region (d), where the count rate is evidently

flat. However, adopting a linear ordinate scale, as in Fig. 7.18(b), suggests operation at a higher voltage. The seemingly flat presentation in Fig. 7.18(a) is actually an illusion created by adopting a log scale on the ordinate, because the plateau characteristic always slopes. The cause of this lies partly in the choice of variables and partly as a consequence of signal-induced background (afterpulses). Slope in the plateau characteristic is clearly a source of uncertainty if attempting absolute measurements. It is shown in Appendix A, Fig. A.11, that $S(>h_0, V)$ approaches the total count as $V \rightarrow \infty$ and is thus unattainable; a better determination of total counts is achieved by extrapolating the integral count rate, $S(>h, V_0)$, using h settings $< h_0$. Examples illustrating this are given in Figs A.3 and A.15 in Appendix A.

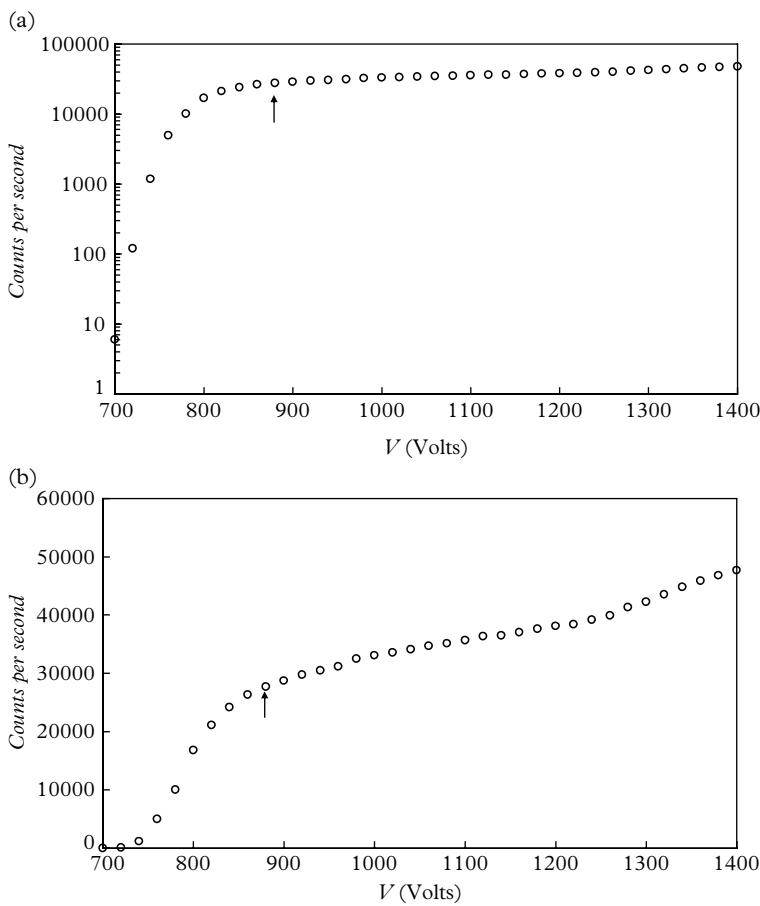


Fig. 7.18. (a) A measured plateau characteristic using a log scale for the ordinate, and (b) the same data plotted on a linear scale. The point to make is the obvious one. The arrows in (a) and (b) refer to the same operating voltage.

The sensitivity of any detector should remain stable with time and temperature, with a degree of independence from small fluctuations in voltage. In photon counting, it is possible to minimize the effect of gain change by choosing the appropriate operating point, while in DC detection it is not. Cooling a PMT reduces dark current and dark counts, but, concurrently, also the long wavelength response. Multiplier gain change per dynode is immeasurably small for individual SbCs and BeCu dynodes. However, it is barely measurable for a set of dynodes; the overall temperature coefficient for a 12-stage PMT is typically $-0.2\%/\text{ }^{\circ}\text{C}$ in the vicinity of room temperature and, consequently, temperature changes have negligible influence in photon counting, although they may be relevant in DC detection. The quality of low-light-level measurements, therefore, is enhanced by cooling and maintaining a fixed temperature, thereby attaining stable QE and gain, while achieving reduced dark counts and dark current. Figure 7.19 shows that there is a significant contribution to dark counts originating from the first dynode. The amplitude of these signals is scaled downwards by the cathode-d₁ gain and, accordingly, also their contribution to dark current. However, small pulses above the threshold contribute to dark count, to the same extent as the larger ones, since all pulses have unity weighting. A dark count plateau characteristic invariably shows more structure than the one for signal. At high light levels, the contribution from background is of

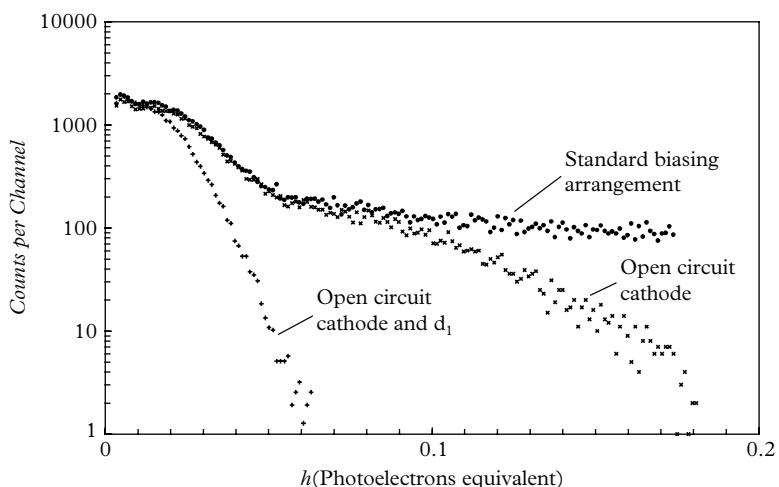


Fig. 7.19. Verification that the majority of small pulses in the background originate from dynodes. The scale on the abscissa has been expanded for this purpose. The voltage-divider biasing arrangement in this example applies three times the inter-dynode voltage between the cathode and d₁, resulting in a d₁ gain of 17; the remaining dynodes each provide a gain of approximately 4.

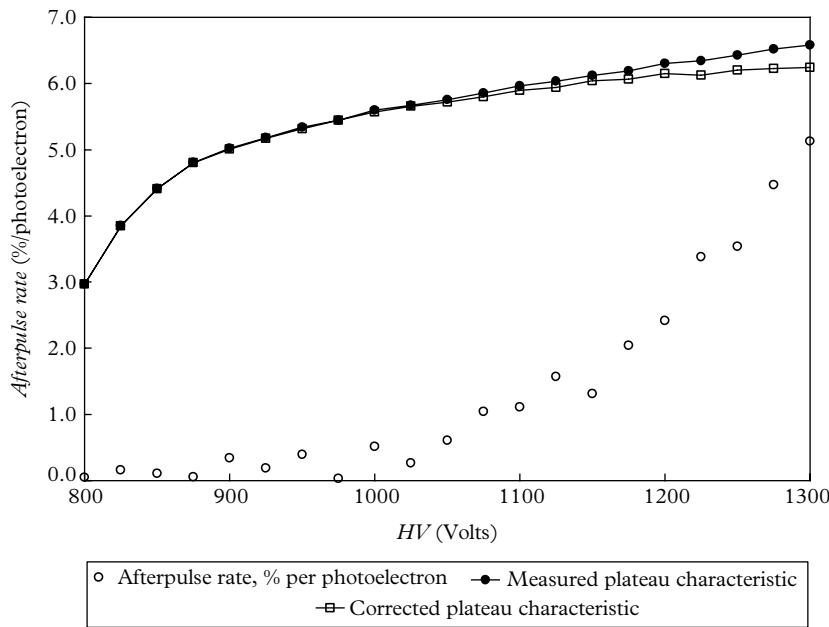


Fig. 7.20. The plateau characteristics for a 9107B (●), the afterpulse rate (○), measured as a function of HV , and a corrected plateau characteristic (□), derived by subtraction of the afterpulse contribution. Afterpulse rates were determined by the variance method described in 4.17.1.

7

limited importance, but may significantly affect the choice of operating point at low light levels.

PMTs designed for photon counting have been optimized to achieve low afterpulse rates (see Fig. 7.20). This refers in particular to the restricted photocathode area types, such as the 9863/100B, but also to the 28 mm 9107B, for example, from ET Enterprises and the HPK R6095P. Multi-photoelectron afterpulses are particularly rare in these PMT types, but those originating between dynodes show up at elevated voltages. Operation at or below 1000 V would be recommended for this particular 9107B because of the build up of afterpulses. It is clear from the differential of the plateau characteristic that such an operating point provides high counting efficiency, with a small contribution from afterpulses. There is a means for reducing the contribution from afterpulses by including dead time in the associated instrumentation. The required dead time need be no more than 5 μ s, for the PMT examined in Fig. 7.21, although at low gain a wider window may be necessary. The benefit of imposed dead time is confirmed in Fig. 7.21.

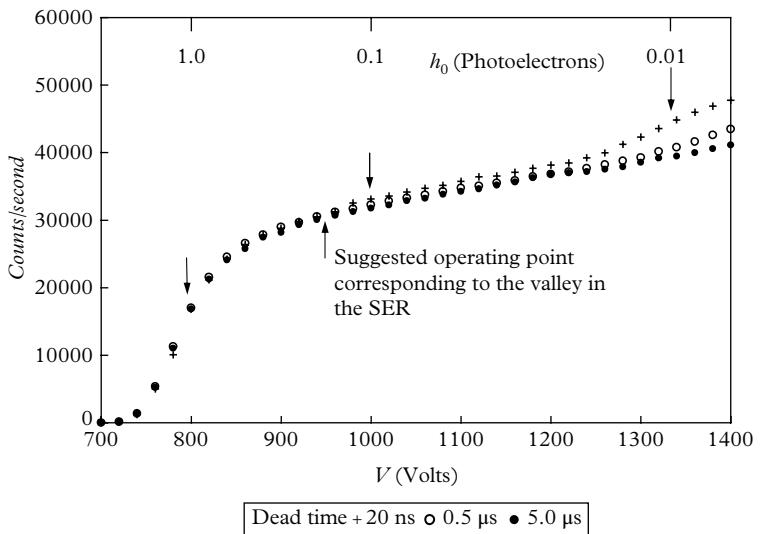


Fig. 7.21. Plateau characteristics measured at fixed h_0 and at three dead time settings.

7

7.6.5 Discriminator threshold

Selecting and setting the trigger threshold is an important and necessary precursor to counting. Given a discriminator of variable threshold and a PMT capable of operating satisfactorily over a wide range in gain, how should the individual sensitivities of the discriminator and the PMT be set to obtain optimal overall performance? The author's experience is that the threshold for a $50\ \Omega$ system should be set in the region of 2 mV. The traditional NIM-based discriminators have a minimum threshold of 50 mV, requiring a fast amplifier of $\times 10$ gain to provide the necessary overall sensitivity. Integrated units, devised by researchers and industrial equipment manufacturers, combine an amplifier and a discriminator and are referred to as amplifier discriminators. Setting the sensitivity much lower, at say 0.2 mV, makes the detector vulnerable to electrical pick-up and consequential false counting. Setting it higher, at, say, 10 mV, may lead to operation of the PMT at excessively high gain with consequential enhanced dark counts. The question regarding threshold optimization is answered by experimentation, yielding results similar to those of Fig. 7.22. Individual plateau curves have been shifted horizontally to coincide at the half-photoelectron level. Faithful replication in the family of curves, after shifting, is indicative of satisfactory performance by the discriminator. There is one exception: specifically, the 0.35 mV setting shows an excess of counts beyond 1000 V. Operating at a threshold of 2.0 mV has no theoretical justification, although there is a long-established practical precedent. Ideally, all photon-counting equipment needs investigation under the environmental conditions that apply, or are likely to apply. There could be

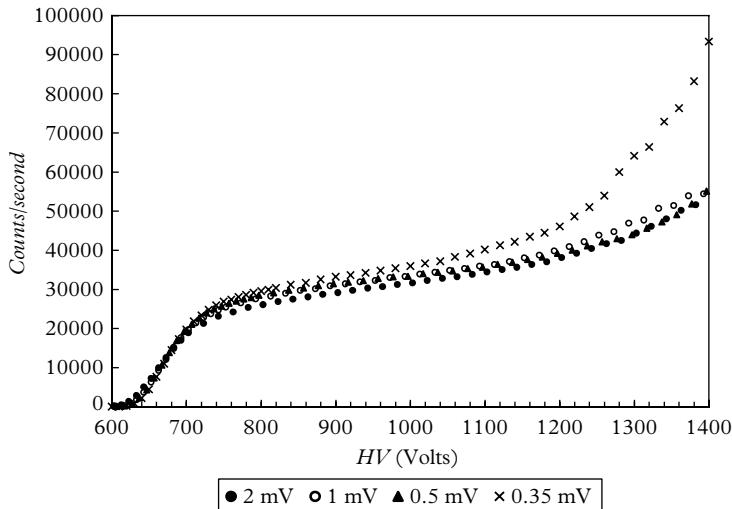


Fig. 7.22. Plateau characteristics measured with a constant intensity light source but at different discriminator thresholds. Curves have been normalized at the half counts by shifting the abscissas. A threshold of 0.35 mV is clearly too low. Each curve has a slightly different count rate at any common plateau voltage—one of the enigmas of photon counting.

7

local contributions from stepper motors and electromechanical switches, as well as from external sources such as electrical equipment and noisy mains. Mains-borne interference is often difficult to detect but the use of an overnight data logger readily uncovers such sources.

7.7 Methods involving DMMs and electrometers

7.7.1 Measurements with a DMM

The schematics in Fig. 7.23 illustrate two basic methods for determining DC anode current by measuring the voltage developed across a load resistor, R_L , of known value. These circuits are suitable for low bandwidth measurements where R_L sets the mean anode voltage and C the bandwidth. The circuit in Fig. 7.23(a) operates at negative HV, with the advantage of maintaining the anode at ground potential. The choice of positive HV, shown in Fig. 7.23(b), sets the photocathode to ground potential, bringing benefits in optical and mechanical interfacing. A battery-operated DMM may be used for either negative or positive HV operation but, in the latter configuration, the voltmeter

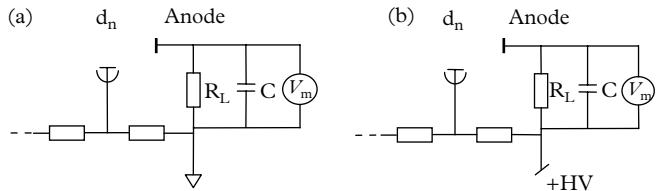


Fig. 7.23. Configurations for making DC measurements: (a) is the traditional way, while in (b) a DMM is floated at HV. Decoupling capacitors have been omitted to aid clarity.

floats at positive HV, posing an electrical hazard. Care must be exercised in configuration shown in Fig. 7.23(b) to avoid damage to instrumentation on switch-on—a low-pass filter or π -network at the positive HV terminal may offer sufficient protection. The HV lead should never be inserted or removed when powered, especially with positive HV.

7.7.2 Electrometers

7

Electrometers are instruments with a history of over 100 years, following the invention of the vacuum triode in 1906. Early models were known as vacuum tube voltmeters, but the products available nowadays utilize junction field effect transistors (JFETs) or metal oxide semiconductor field effect transistors (MOSFETs) to provide devices with very high input impedance of up to $10^{16} \Omega$. This means that they can pick off a voltage at the picovolt level without loading the circuit, and currents as low as 10^{-17} A are possible to measure. Instruments with this sensitivity are at the forefront of detection: handling them requires skill, know-how, perseverance, and, ideally, years of experience. *Low Level Measurements Handbook* (7th ed.), produced by Keithley Instruments (Keithley Instruments Inc. 1972) is particularly useful to anyone contemplating light measurements of this nature. This is essential prior reading when planning to make ultra-low-level measurements.

Electrometers incorporating a transimpedance amplifier at the input are ideal for making current measurements (Fig. 7.24(a)). They are also referred to as current-to-voltage converters—an informative description of what they actually do. A PMT is a current generator and therefore ideally suited for use with such an amplifier configuration of ultra-low-input impedance. Another desirable feature of low-input impedance is the prevention of anode cut-off, as can happen with the configurations shown in Fig. 7.24(a) and (b). Furthermore, low-impedance circuits are less liable to noise pick-up. Electrometers with femtoamp sensitivity are readily available, but the extent of their capability is seldom fully exploited: PMT dark current and leakage, not the amplifier, invariably set the level of detectivity.

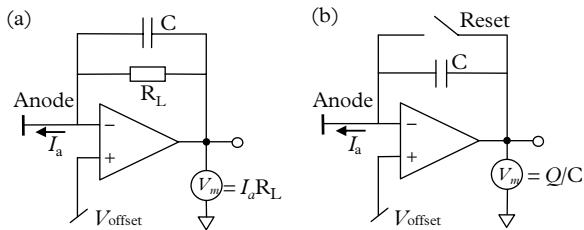


Fig. 7.24. Conceptual front-end configurations used by instrument manufacturers with negative HV. The configuration in (a) performs current-to-voltage transformation with V_m positive. The R_LC combination sets the bandwidth.

7.7.3 Voltage dividers

A uniform divider network with common-value inter-dynode resistors R is the usual choice for low-level-light measurements. The manufacturer's recommendation on biasing $k - d_1$ should be followed: either using multiple resistors or using zener diodes. It follows from the power law relationship between gain and voltage V that

$$\frac{dg}{g} = \beta \frac{dV}{V} = \sim 1\%/\text{Volt},$$

at $V = 1000$ V, and $\beta = 10$.

Gain, and therefore signal output, is sensitive to the stability of the supply where DC detection is adopted. A change of the voltage by 1 V changes the gain by 1 %. But, with reference to Fig. 7.17, count rate only changes by $\sim 0.1\%/\text{Volt}$ when photon counting at $V = 1050$ volts. Modest decoupling, typically 1 nF on the last three stages, is sufficient to deal with the consequences of closely bunched events. Voltage divider standing current can be set lower for photon counting than for DC applications. This benefit derives from operating on the flat region of the plateau characteristic, discussed in 7.6.3. Battery-operated luminometers, for example, use 1 to 10 M Ω resistors in an all-resistor divider; or, an active divider may be chosen with biasing and ballast resistors of up to 30 M Ω in value.

7

REFERENCES

- Arecchi, F. T., Gatti, E., and Sona, A. (1966). Measurement of low light intensities by synchronous single photon counting. *Rev. Sci. Instr.*, **37**, 942–8.
 Bevington, P. R. and Robinson, D. K. (1992). *Data Reduction and Error Analysis for the Physical Sciences* (2nd ed.). McGraw-Hill, New York.
 Foord *et al.* (1969). The use of photomultiplier tubes for photon counting. *Appl. Opt.*, **8**, 1975–89.

- Janesick, J. R., Klassen, K. P., and Elliott, T. (1987). Charge-coupled-device charge-collection efficiency and the photon-transfer technique. *Opt. Eng.*, **26**, 972–80.
- Jones, R., Oliver, C. J., and Pike, E. R. (1971). Experimental and theoretical comparison of photon-counting and current measurements of light intensity. *Appl. Opt.*, **10**, 1673–80.
- Keithley Instruments Inc (1972). *Low Level Measurements Handbook* (7th ed.). Tektronix, Beaverton, OR. Available at http://www.tek.com/sites/tek.com/files/media/document/resources/LowLevelHandbook_7Ed.pdf (accessed 29 Jan 2017).
- Knoll, G. F. (1989). *Radiation Detection and Measurement* (2nd ed.). John Wiley & Sons, Inc., Hoboken, NJ.
- Mandel, L. (1959). Fluctuation of photon beams: The distribution of photoelectrons. *Proc. Phys. Soc.*, **74**, 233–43.
- Oliver, C. J. (1977). *A Comparison of Current Measurement with Photon Counting in the Use of Photomultiplier Tubes*. ET Enterprises Reprint Series, RP066. ET Enterprises Ltd, Uxbridge.
- Oliver, C. J. and Pike, E. R. (1968). Measurement of low light flux by photon counting. *J. Phys. D: Appl. Phys.*, **1**, 1459–68.
- Pao, Y. H. and Griffiths, J. E. (1967). Performance of a new photoelectric detection method for high-resolution optical spectroscopy. I Investigation of basic characteristics. *J. Chem. Phys.*, **46**, 1671–9.
- Saleh, B. E. A. (1973). Photon time of arrival, time between consecutive photons and the moment generating function. *J. Phys. A: Math., Nucl. Gen.*, **6**, L161–L164.
- Young, A. T. (1969). Photometric error analysis. IX. Optimum use of photomultipliers. *Appl. Opt.* **8**, 2431–47.

8

Timing with PMTs

8.1 Electron motion in electric fields

8.1.1 Introduction

PMT-resolving time refers to the precision with which the PMT can measure the time interval between events. Such events may be single- or multi-photon in nature, and may involve the difference in time between an electrical time marker, such as a laser trigger pulse, and subsequent photon detection. Increasingly in the astrophysics field, for example, researchers need an absolute timestamp to search for coincidences with other types of detector, possibly sited in distant locations. By events, we generally refer to signal detection at the anode, although as an intermediate step we sometimes concern ourselves with the statistics of photodetection events at the photocathode. The ability of a PMT to distinguish two closely spaced photo detections is allied to, but is not the same as, its timing fidelity; the former is relevant to the topic of dead time in photon counting, which is covered in 4.19.

Several contributions to resolving time originate within a PMT. A fundamental timing specification is quoted with reference to the detection of single photons from a synchronous source, and is relevant to both the development of PMTs and the selection of a suitable device by the user. In applications involving multi-photon pulses, the combined statistical performance of a PMT and a light source needs to be understood, including the statistics of photoemission. The time sequence for the progression of single-photon-initiated signals through a PMT is as follows:

8

- (1) the arrival of a photon at the photocathode
- (2) the generation and release of a photoelectron at the photocathode
- (3) the arrival of a photoelectron at the input to the multiplier
- (4) the delivery of the charge pulse at the output of the multiplier
- (5) the collection of a charge pulse at the anode

Photoelectron emission is not instantaneous because of the finite time required by an electron to escape the photocathode. The classical model of photoemission predicts that this lag, together with its variance, is of the order of picoseconds. Experimental data verify that the delay increases with wavelength, approaching 100 ps at a wavelength of 860 nm (Bazarov *et al.* 2008). Only the fastest, discrete dynode PMTs are capable of sensing this small delay, but it is readily observed with MCP devices. This source of time dispersion is generally unimportant in conventional PMTs.

In end window devices, the photocathode is laid down on a plano-planar, plano-concave, or plano-prismatic window and, in the plano-concave construction, the first dynode is generally located at the radius of curvature of the window, thus equalizing path lengths between all pairs of points on the cathode and the first dynode; this is only strictly achieved for those photoelectrons emitted normal to

the surface and with zero initial velocity. PMTs intended for slow applications incorporate a less costly plano-planar window; photoelectrons from the periphery take up to a nanosecond longer to reach the first dynode than do those emanating along the axis of the window. This leads to considerable dispersion but it is of no concern if the intended applications of such PMTs are for spectroscopy, for example.

Every photoelectron follows a unique path to the first dynode with a corresponding transit time. Photoelectrons released from the same point on the photocathode follow different paths to the first dynode, since every photoelectron is subject to two statistical variables, namely, launch angle and emission energy. In the multiplier, each dynode, of which there may be up to 19, contributes to dispersion through this process. The way in which the output charge is collected by the anode also makes a contribution to time dispersion. We will consider these sources of time dispersion in photoemission and in secondary emission.

It is convenient to express photon energy in electron volts, using the well-known relationship: $E = hc/\lambda$. Wavelengths of 400, 600, and 1000 nm, for example, are equivalent in energy terms to 3.1, 2.1, and 1.2 eV, respectively. Photoelectrons always carry away less than the original photon energy, in accordance with an energy spectrum of complex shape (Apker *et al.* 1953; Nathan and Mee 1970; Spicer 1972). As a generalization, for photons in the range $300 < \lambda < 800$ nm, the maximum photoelectron energy is about 1 eV less than the photon energy. For blue light, and a bialkali cathode, the initial energy, V_0 , ranges from zero to ~ 2 eV. The sources that give rise to dispersion in the multiplication process (secondary emission) are similar to those for photoelectron emission. But in the multiplication process the spectrum of initial energies has a higher mean in the region of 5 eV, and the distribution in launch angles is understood to be Lambertian.

The transit time, T, of a PMT is defined as the time between the generation of a single photoelectron at the photocathode and the appearance of a charge pulse at the anode. It is obvious that T can only be measured if there is a means of signalling the time at which the photon was generated, correcting for any time of flight between the light source and the photocathode—if necessary. An electrical pulse, or edge, synchronous with a light flash is required to define the start, and the same electrical pulse that is used to trigger the light source may serve this purpose. The source for a short duration pulse of light is typically a laser-based system or a light-emitting diode (such as the XP22) excited by an avalanche transistor (see 14.12). The trigger pulse for both these light emitters also provides an electrical signal coincident with a multi-photon light flash. Laser-based light sources are capable of producing pulses of light of very short duration—measured in femtoseconds—and the pulse width for some LEDs is as low as 100 ps.

PMTs are designed nowadays using 3-D electron-optical simulation programs marketed under names such as ‘Maxwell’ and ‘Simeon’. The paper by Tahir *et al.* (1991) describes the electron-optical simulation of front ends and dynode structures for certain PMTs manufactured by Electron Tubes; this paper is particularly readable for those new to the subject. The way in which the simulation is done is

as follows: the user provides the mechanical layout of the electrodes, and the applied voltages from which the potential distribution, $V(x, y, z)$, is generated are determined by solving the Laplace equation. Ray tracing establishes the landing position and the transit time from any point on the photocathode to the first dynode. By taking a set of starting positions distributed across the entire photocathode area, the user can decide whether a proposed design meets given timing, or efficient collection, requirements. An important point to make here is that the computer program is a simulation only, without original design capability. This is where the insight and experience of the programmer is essential, since an electron-optical program only provides verification of a proposed design—it is a powerful tool, but that is all.

This chapter begins with an elementary discussion on the laws of electron motion in an electric field, starting from the equations of Newton. This, after all, is the very basis on which PMTs operate, and the timing performance of real devices can be inferred from basic laws of motion. Furthermore, the relationship of time dispersion to field strength, initial electron energy, angle of emission, and electrode spacing follows from these laws of motion; such an analysis provides a means for interpreting the results of computer simulation and actual measurements of PMT performance. For conventional PMTs, the major contribution to dispersion arises from the transition between the cathode and the first dynode. It will be shown that the attainment of fast timing depends primarily on the field gradient at the photocathode: the higher the magnitude of this parameter, the better the timing. The next section covers the contribution to timing made by the multiplier, culminating in review of the timing attainable from the traditional PMTs with discrete linear focussed dynodes, and the exceptional performance attainable from the devices of more recent design. The functional dependence on time, expressed as a current pulse, $i(t)$, for single-photon excitation is the basis for analysis. Multi-photoelectron pulses of any waveform, in particular, the light output from a scintillator, may be represented by a superposition of many versions of $i(t)$, each one suitably delayed and scaled in amplitude. It will be shown that Monte Carlo simulations of PMT output pulse shapes compare favourably with actual oscilloscope measurements, and the random waveforms thus generated can be used to investigate the performance of electronic circuits intended for fast timing. The preferred technique for obtaining minimal time dispersion is that of CF discrimination, which is discussed in detail. As with most PMT applications, statistical concepts and considerations lie at the heart of performance, setting ultimate achievable limits. Timing artefacts, caused by spurious PMT behaviour, are discussed in 11.5.

A charged particle in a uniform electric field E experiences a constant force in the direction of increasing potential, V , and by definition dV/ds is constant. The equations that apply under these conditions are the same as those for a body falling under the influence of gravity. The speed of electrons and ions, in the context of PMTs, are always many orders of magnitude lower than the speed of light so that relativistic effects may be ignored. Electric fields in vacuum devices are generally non-uniform, with the exception of proximity detectors. The analysis that follows

considers linear and non-linear field gradients, some of which are non-analytic. In addition, the present analysis is confined to two dimensions. Nevertheless, it can be shown using 3-D, computer-aided, electron-optical simulations that the predictions of classical Newtonian laws of motion (2-D only) are both relevant and useful. Formulae which follow from the Newtonian analysis predict the dependence of time dispersion on distance, s_0 ; electric potential, $V(s_0)$; initial energy, V_0 ; and angle of emission, θ . In general, the field distributions are non-uniform, and the special case of square law dependence is investigated, enabling a comparison to be made with linear and actual PMT distributions. Certain PMTs conform to a linear (proximity focussed) distribution, while others follow a square law dependence (Fig. 8.1).

A grasp of the basic physics describing the motion of electrons in PMTs is valuable both to the designer and to the user of PMTs. The relationships between the variables that follow from these fundamental considerations serve both as a guide to PMT design and as an aid to realizing the full capability of a particular device. The interest of the user lies in realizing the best performance from a given device, and the only way open to optimizing timing is through tuning the voltage distribution of the divider; it is quite remarkable the degree to which performance can be optimized through judicious tuning of the applied voltages. The term 'jitter' is a general term frequently used to indicate that a timing parameter, such as transit time, is subject to statistical fluctuation. There are several ways in which to quantify a given phenomenon statistically: standard deviation, variance, and fwhm. The choice usually depends on convention.

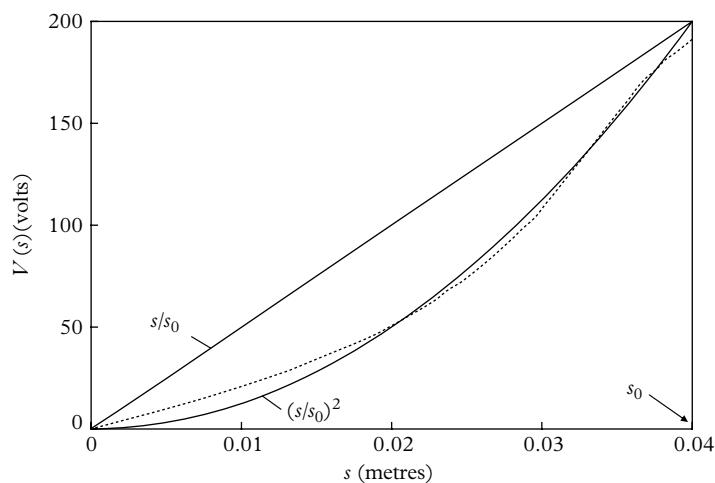


Fig. 8.1. The dashed curve represents the potential distribution between photocathode and first dynode, derived from an electron-optics simulation for a 9814B with $k-d_1$ spacing of 40 mm (see Fig. 8.7). The uppermost curve is for a linear voltage distribution, and the lower, solid one, describes a square law, $V(s) = V(s_0)(s/s_0)^2$.

From Newton's second law,

$$m \frac{d^2 s}{dt^2} = -Ee \quad (8.1)$$

where:

m is the mass of the particle: 9.1×10^{-31} kg for an electron, and 1.67×10^{-27} for a proton, H^+ , for example

s is the distance in metres

$E = -dV/ds$ is the potential gradient in volts per metre

e is the electronic charge

According to the laws of electron motion in an electric field, the work done in moving a charge from Point A to Point B , W_{AB} , is given by the difference in potential between these two points:

$$W_{AB} = e(V_B - V_A). \quad (8.2)$$

We equate the change in potential energy to a gain in kinetic energy; thus,

$$\frac{1}{2}m(v_B^2 - v_A^2) = e(V_B - V_A).$$

For presentation reasons, we replace v_B and V_B by v and V , respectively, and v_A and V_A by v_0 and V_0 , respectively, obtaining

$$\frac{1}{2}m(v^2 - v_0^2) = e(V - V_0). \quad (8.3)$$

Equation (8.3) gives the speed in terms of the field potential. It is quite general and without restriction to uniform or to one-dimensional fields. This is because (8.3) expresses conservation of energy and hence it is independent of the electron trajectory. Consider an electron with initial velocity v_0 at the origin of a reference frame in which $s=0$. From (8.3) we can express this velocity in terms of a potential, V_0 ; thus,

$$v_0 = (2e/m)^{1/2} V_0^{1/2} = 5.93 \times 10^5 V_0^{1/2} \text{ m/s}, \quad (8.4)$$

noting that 1 eV corresponds to a velocity of 5.93×10^5 m/s. The velocity of the electron at any point s in a field, $V(s)$, is

$$v^2 = v_0^2 + (2e/m)V(s).$$

Substituting for v_0 by using (8.4), we obtain

$$v^2 = (2e/m)(V_0 + V(s)). \quad (8.5)$$

Note that the final velocity is independent of both the path taken and the functional form of $V(s)$. The velocity of an electron in a field distribution given by $V(s) = V(s_0)s/s_0$ is shown in the upper set of curves in Fig. 8.2; the lower set of curves is for $V(s) = V(s_0)(s/s_0)^2$, and relates to 8.1.4. It is obvious in all the curves of Fig. 8.2 that the time to cover a unit distance is greatest when the velocity is low: that is, at the start of motion.

We note for the non-uniform field, which is described by a square law, that the velocity remains relatively constant for the first few millimetres of path, while an electron in a linear field experiences rapid acceleration immediately after emission. Since the time taken per unit increment in distance varies as the reciprocal of the velocity, clearly the major contribution to dispersion arises from the early track history of each electron. We investigate the motion of an electron in three different field configurations: in the first, the potential varies linearly with s ; in the second, it is assumed to vary as s^2 ; and in the third, the field distribution for an actual PMT is explored. Such distributions are representative of practical devices, and analytical solutions can be found in the case of the first two. The methodology for handling any measured distribution $V(s)$ is one of numerical integration, using the field pattern of a PMT, as determined by electron-optical simulation (see 8.1.3, Fig. 8.4, and 8.2.1, Fig. 8.7).

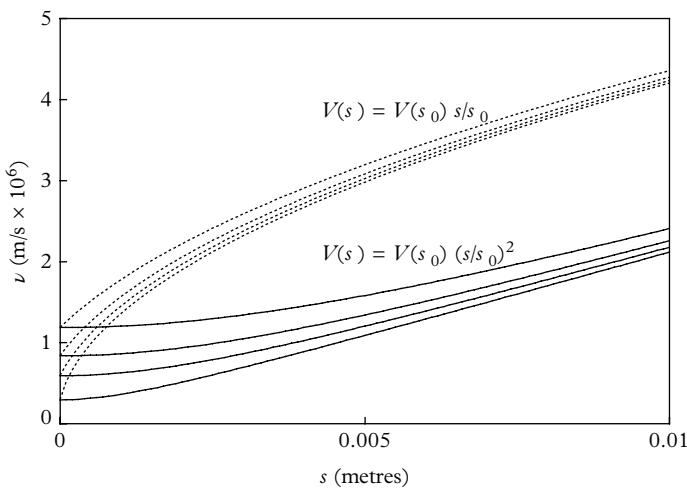


Fig. 8.2. Electron velocity, calculated from (8.5) as a function of distance from the origin, for initial energies (velocities) of 4.0, 2.0, 1.0, and 0.25 eV. The uppermost curve in each set corresponds to 4.0 eV; $V(s_0)$ is 200 V, and s_0 is 40 mm but only the initial 10 mm path of each curve is shown.

8.1.2 Electron motion in a uniform electric field

The time taken to move from $s = 0$ to $s = s_0$ in a uniform field defined by $V(s) = V(s_0)$ s/s_0 follows from (8.5):

$$t = \int_0^{s_0} \frac{ds}{v} = \frac{2s_0}{(2e/m)^{1/2} V(s_0)} \left[\left(V_0 + V(s_0) \right)^{1/2} - V_0^{1/2} \right]. \quad (8.6)$$

If $v_0 = 0$ or, equivalently, $V_0 = 0$, (8.6) reduces to

$$t_0 = \frac{(2m/e)^{1/2}}{\left(V(s_0) \right)^{1/2}} s_0. \quad (8.7)$$

Emission energies, V_0 , range up to about 3 eV for photoelectrons produced by visible light, and applied voltages are typically within 100 to 500 V, depending on the PMT diameter. Hence, (8.7) is a good approximation, within a few per cent, for typical values of V_0 . The order of magnitude of the $k-d_1$ transit time for PMTs of diameter ~ 50 mm follows from (8.7) by noting that $(2m/e)^{1/2} = 3.4 \times 10^{-6}$, that $V(s_0)$ is typically 200 V, and that s_0 is of the order of 50 mm, leading to t_0 of about 10 ns. Note that the transit time varies as $s_0/(V(s_0))^{1/2}$ and, if $V_0 \ll V(s_0)$, we may take $(V_0 + V(s_0))^{1/2} = (V(s_0))^{1/2}$ in (8.6), giving

$$\begin{aligned} t &= \frac{(2m/e)^{1/2}}{\left(V(s_0) \right)^{1/2}} s_0 - \frac{(2m/e)^{1/2}}{V(s_0)} s_0 V_0^{1/2} \\ &= t_0 - \Delta t, \end{aligned} \quad (8.8)$$

8

where t_0 refers to $V_0=0$, being the transit time for an electron of zero initial velocity, and Δt is a correction to the transit time that allows for an initial energy ‘kick’ in the emission process, accepting for the moment that $0 < V_0 < 3$ eV for most photocathode materials. According to (8.8), the transit time varies as $1/(V(s_0))^{1/2}$, for fixed s_0 and V_0 , whereas the jitter, Δt , varies as $1/V(s_0)$, with an increase in the applied voltage leading to a reduction in jitter. Note that transit time itself poses no loss in timing performance: it is the variation in this parameter, from one event to the next, that causes dispersion. The major contribution to jitter comes from the initial part of the trajectory, where the electron is slow moving, and hence it is the strength of the extraction field at the cathode that is crucial. Equation (8.8) can be expressed as

$$\Delta t = (V_0/V(s_0))^{1/2} t_0, \quad (8.9)$$

which is also given by Beck (1976). We note, by substituting typical values for V_0 and $V(s_0)$ in (8.9), that Δt can represent as much as 10% of t_0 .

We can deduce the order of magnitude of the transit-time dispersion with wavelength from Fig. 8.3. Examining the curve close to the origin indicates $\Delta t \approx 1$ ns/eV of initial energy, and recall from (2.9) that, at 1 eV ≈ 1240 nm, we

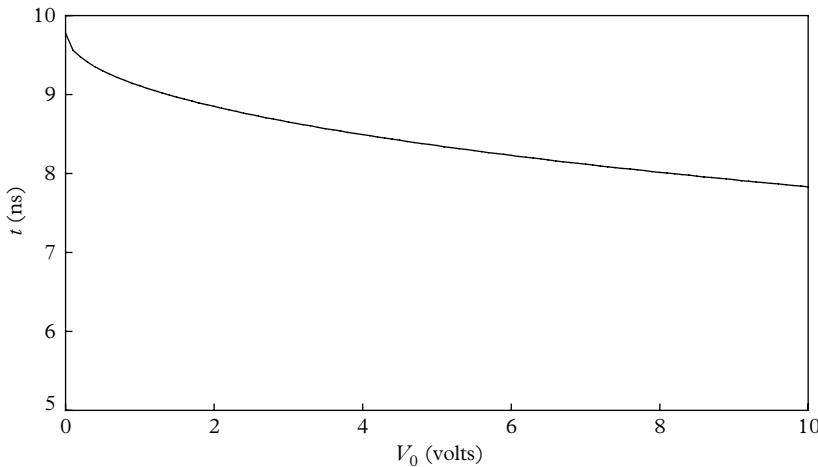


Fig. 8.3. Transit time, t , as a function of V_0 using (8.6) or (8.8), with initial velocities, expressed in electron volts. The curve is for $V(s_0) = 200$ V, and $s_0 = 0.040$ m. Typical emission velocities of photoelectrons range from zero to a few electron volts, with consequential wavelength dependent time jitter of up to 1 ns.

have $\Delta t \approx 1$ ps/nm of wavelength. This is consistent with measurements quoted by Hungerford and Birch (1996) for Hamamatsu R928 (side window) and Philips XP2020; both exhibit a dispersion of 0.5 ps/nm.

8

8.1.3 Electron motion in non-uniform electric fields

In proximity focussed devices, such as an MCPPMT, the electric field between the cathode and the MCP is linear and uniform over the effective photosensitive area. Discrete dynode PMTs, such as the 9814B, however, include focussing elements that produce a field profile approximating to a square law (Fig. 8.1). Using (8.4) and (8.5), we now have

$$\begin{aligned} v^2 &= v_0^2 + (2e/m) V(s) = (2e/m) (V_0 + V(s_0)s^2/s_0^2), \\ t &= \int_0^{s_0} \frac{ds}{v} = \frac{(m/2e)^{1/2}}{\left(V(s_0)\right)^{1/2}} s_0 \int ds / (V_0 s_0^2/V(s_0) + s^2)^{1/2} \\ &= \frac{(2m/e)^{1/2}}{2\left(V(s_0)\right)^{1/2}} s_0 \left[\ln(s_0 + (s_0^2 + s_0^2 V_0/V(s_0))^{1/2}) - \ln(s_0^2 V_0/V(s_0))^{1/2} \right] \end{aligned} \quad (8.10)$$

and, since $V_0/V(s_0) \ll 1$, (8.10) reduces to

$$t = \frac{(2m/e)^{1/2}}{4\left(V(s_0)\right)^{1/2}} s_0 \left(\ln(4V(s_0)/V_0) \right). \quad (8.11)$$

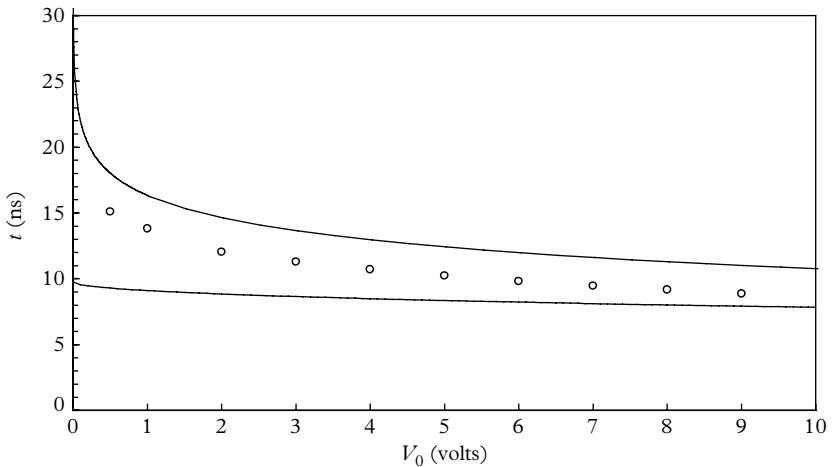


Fig. 8.4. Comparison of the timing performance for $V(s) \approx s^2$ (uppermost curve), based on (8.11), with that for a uniform field, $V(s) \approx s$ (lowermost curve), in accordance with (8.8). The data represented by circles follows from numerical integration, $\int ds/v$, of the potential distribution for a 9814B PMT, as given in Fig. 8.7. Assumed parameters are $V(s_0) = 200$ V, and $s_0 = 40$ mm.

8

For low values of V_0 , there is a sharp turn-up in transit time and consequently in jitter (upper curve of Fig. 8.4). This has already been explained in terms of the slow initial acceleration of an electron following emission from the cathode. Low acceleration is to be understood in terms of the low electric field, $E(s) = -dV(s)/ds$, in the vicinity of the starting position of the electron. Note that $E(s) = -2V(s_0)s/s_0^2$, and $E(0) = 0$; the electron needs initial energy (velocity) if it is to move from its rest position. However, it is clear, because of the sluggish field gradient in the vicinity of the photocathode, that the jitter will be considerable for all starting energies below a few electron volts, as illustrated. On the other hand, the field gradient for a uniform field is $-V(s_0)/s_0$, and a ‘kick’ is not required to initiate motion.

8.1.4 Laws of motion applied to PMTs

Equations (8.9) and (8.11) predict the order of magnitude of the chromatic contribution to jitter. They also indicate that jitter may be reduced by increasing the applied voltage and decreasing electrode spacing, s_0 . Consideration has been given so far to the effect that V_0 has on timing, and no allowance has been made for paths longer than s_0 . Cathode-to-first-dynode path lengths, in a PMT with a plano-planar window, vary significantly from centre to edge—up to 10 mm. Fast PMTs invariably incorporate a plano-concave window with a radius of curvature equal to the $k-d_1$ spacing, s_0 . But, even with this type of window, there will be paths of different lengths because of the angular distribution of the

photoelectrons. The additional transit time incurred for a small extension to s_0 follows from (8.7); thus,

$$\Delta t = \frac{(2m/e)^{1/2}}{V(s_0)^{1/2}} \Delta s_0. \quad (8.12)$$

Inserting $V(s_0) = 200$ V, we have $\Delta t = 240$ ps/mm but note that Δt is independent of s_0 . For moving across the photocathode, measured values quoted by Hungerford and Birch (1996) range over 250 to 450 ps/mm for side window R928 and R955 PMTs. This path-length dependency poses a serious limitation on the otherwise fast response of side window PMTs and they are sometimes stopped-down for fast applications. The recognized standard for fast applications, the XP2020 family, has no such shortcoming, giving greatly improved performance in this respect: the time dispersion moving away from the centre towards the periphery of the PMT window is on average about 12 ps/mm, although, within 1 cm of the axis of the PMT, the spatial dependency is half this value. (see Piétri 1981, Fig. 4.6, and Philips 1994, Fig. 4.7).

There is an important dependency that applies to all electron-optical configurations, regardless of their complexity: an increase in the overall voltage leads to a reduction in Δt ; (8.9) to (8.11) bear this out. If timing is the major consideration, $V(s_0)$ should be set high without compromising other performance parameters, such as SER or dark counts. Obviously, if the signal can be concentrated within a reduced diameter of the photocathode, the range in path lengths, and consequently the jitter, will be limited, in accordance with (8.12). The formulae presented thus far benefit designers and users of PMTs; they also explain the evolution of fast and ultra-fast PMTs.

8.1.5 Spot size

We have studied the effect of initial energy on PMT timing but there is a further aberration to consider. If the y coordinate lies along the axis of the PMT, the time taken to reach s_0 is given by (8.7). If the initial velocity is parallel to the y -axis, there is no displacement. However, an electron launched in the plane of the photocathode, (90°), suffers the maximum displacement from the y -axis. Such an electron moves with constant velocity in the x direction for a time t , with a displacement r_{\max} , given by the product of (8.4) and (8.7):

$$\begin{aligned} r_{\max} &= v_0 \times t \\ &= (2e/m)^{1/2} V_0^{1/2} \times (2m/e)^{1/2} s_0 / V(s_0)^{1/2} \\ &= 2s_0(V_0/V(s_0))^{1/2}. \end{aligned} \quad (8.13)$$

The displacement for photoelectrons emitted at an angle to the normal is given by (8.13), after scaling by $\sin\theta$. Electrons follow parabolic paths, landing a distance r off axis, as shown in Fig. 8.5. Figure 8.6 presents a family of electron trajectories covering a range of launch angles.

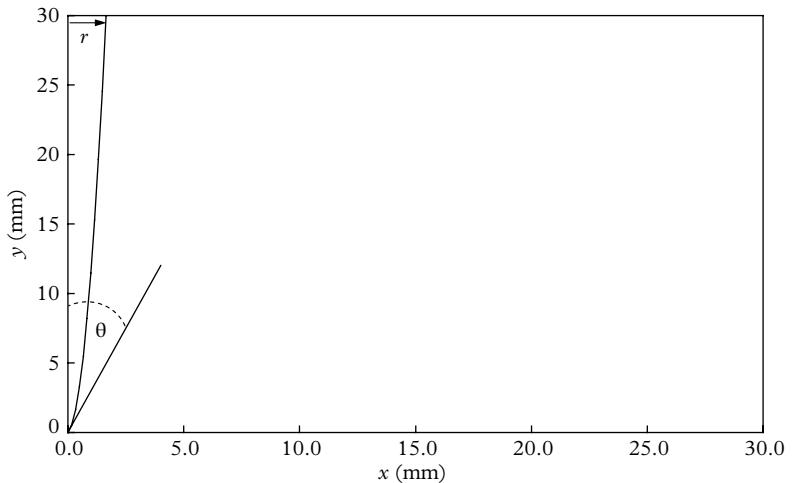


Fig. 8.5. A photoelectron with initial energy and a launch angle θ suffers a displacement r at the first dynode. The path is parabolic, and the displacement is relatively small compared with s_0 .

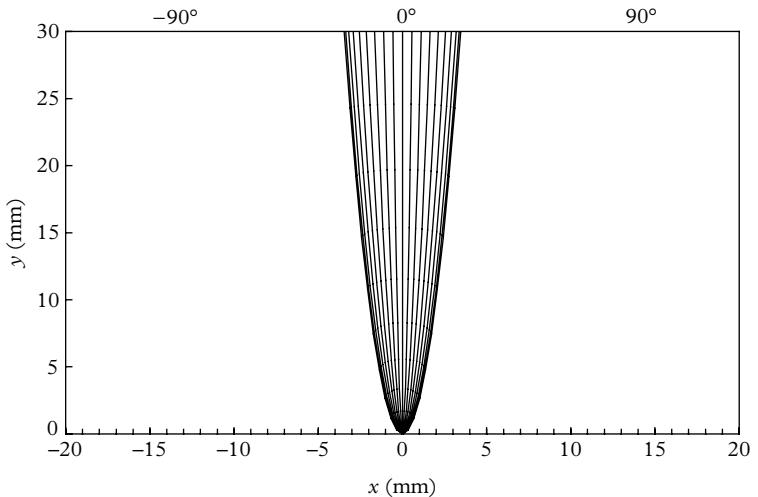


Fig. 8.6. Electron trajectories for $V_0 = 1$ eV, and launch angles ranging from 0 to 90° in steps of 10° ; $V(s_0) = 300$ V, and $s_0 = 30$ mm. From (8.13), the maximum displacement is ± 3.46 mm.

The contribution to time dispersion, because of finite spot size on the receiving electrode (assumed to be d_1), arises from the distribution of starting positions on this electrode. Linear focussed dynodes are particularly asymmetrical in this regard. The equation for r_{\max} predicts a spot size varying linearly with s_0 and as

the inverse square root of the applied voltage, on the assumption of a linear voltage distribution. The same procedure may be followed to determine r for a non-uniform voltage distribution, using (8.11). Figure 8.4 predicts a considerable reduction in transit time from the action of V_0 .

The spot size for the parameter values assumed here is possibly normal, with an fwhm in the range of 2 to 5 mm. Equation (8.13) highlights the well-known problem with large-diameter PMTs about the required area of d_1 . The $k-d_1$ distance, s_0 , is typically 100 mm in a 5" PMT and 30 mm for a 2" PMT and, at the same applied voltage, $V(s_0)$, the spot size increases by a factor of 100/30 over that for the example illustrated in Fig. 8.6. To keep r_{\max} the same for both tube types would require V_{k-d_1} some ten times higher in the 5" PMT because of the square root dependence in (8.13). This requirement poses practical problems for designers and manufacturers of 5" PMTs. The recommendation for large-diameter PMTs such as the XP2041, RCA8854, 9823B, and H6527 is in the region of 1000 V for the $k-d_1$ voltage. All large-diameter PMTs, apart from other considerations, suffer loss in collection efficiency because of spot size considerations. The formulae presented predict only an order of magnitude for time dispersions illustrated in Fig. 8.4, for example. Improved predictions based on a three-dimensional statistical approach are needed. This is complicated because of the number of statistical processes, and the multiplicity of sources (dynodes) that have to be included. Some voltage divider configurations include a potentiometer between the first and third dynode, with the centre tap connected to the second dynode for fine tuning (discussed further in 8.2.2).

8.2 Evolution of fast linear focussed PMTs

8.2.1 The PMT front end

Fast PMTs, such as the EMI 9814B and the RCA 8850, available in the 1960s, included a single focus electrode disposed between the cathode and the aperture plate of the first dynode. The recommendation is to maintain the focus ring at the potential of d_1 to give both good timing and acceptable collection efficiency, although the primary purpose of the ring is gating (see 13.9). The front-end design of Fig. 8.7 produces the quasi-square law distribution shown in Fig. 8.1.

In the late 1960s, Philips, known in the US as Amperex, introduced a family of ultra-fast PMTs with diameters of 5", 3", and 2". The PMT with the latter diameter, the XP2020, set the standard for 40 years as the preferred PMT for fast-timing applications. The innovative step taken by Philips with the XP2020 was the addition of a second electrode, as shown in Fig. 8.8. An accelerator, located between the cathode and d_1 , is internally connected to d_5 , thereby operating at typically 1000 V with respect to the photocathode. The boost to the field

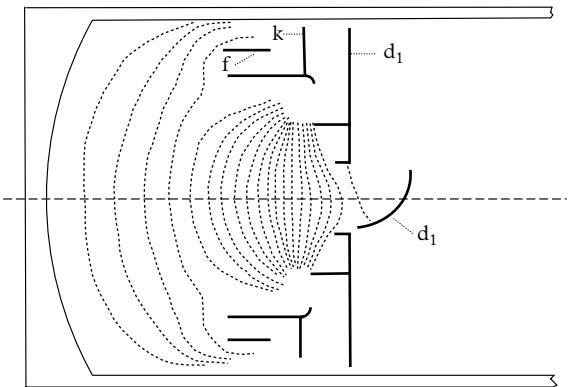


Fig. 8.7. Equipotentials, shown for a 9814B fast PMT, are spaced at 12.5 V intervals. Extraction field strength is 2.5 V/mm for $V_{k-d_1} = 250$ V.

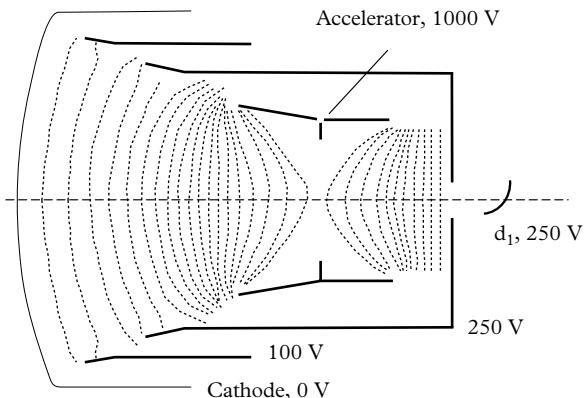


Fig. 8.8. The field strength at the photocathode of an XP2020 is primarily determined by the accelerator voltage; at 1000 V, this produces an extraction field of 14 V/mm. The voltage contours are 50 V apart in this representation.

gradient at the photocathode is shown in Fig. 8.9. Note that photoelectrons actually decelerate in the region between the accelerator and the first dynode, but this has little effect on timing since velocities are substantial at this stage. The more recent Hamamatsu R2083, an eight-stage PMT, incorporates an accelerator operated at a potential between d_6 and d_7 , and it is consequently also very fast.

The relationship between jitter and the photocathode extraction field can be verified experimentally by studying the performance of a range of manufactured PMTs (see Tables 8.1 and 8.2 and Fig. 8.10). The trend is clear, although the scatter in the individual results points to the existence of other sources that reduce timing

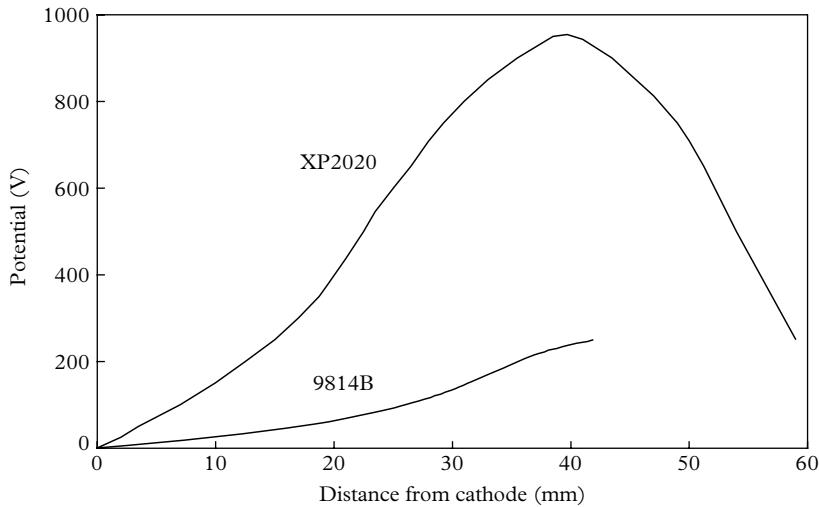


Fig. 8.9. The field gradient in the cathode to first dynode region of two fast PMTs; V_{k-d1} is 250 V in both cases but the boosting effect of the accelerator electrode is clearly evident in the XP2020. The extraction field at the cathode, upper curve, is 14 V/mm while, for the 9814B, it is 2.5 V/mm at $V_{k-d1} = 250$ V.

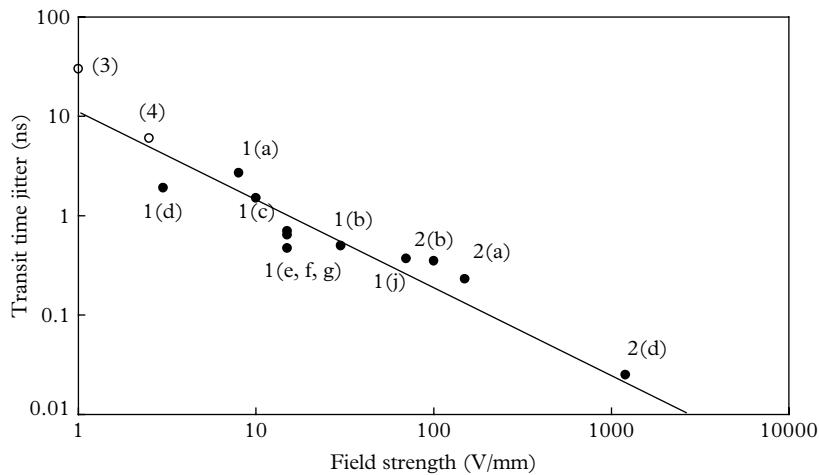


Fig. 8.10. Illustrating the relationship between the field strength at the photocathode and the transit-time jitter (fwhm) for a range of PMTs of different types. The labels link the points to the entries in Tables 8.1 and 8.2. Open circles represent PMT's intended for spectroscopy applications: (3) is for a 30 mm box-and-grid PMT, and (4) is for a 50 mm, plano-planar window, with a T-cup type first dynode. The trend of the PMT data is to follow a linear relationship between field strength and jitter, but the scatter indicates there are other sources of time dispersion.

Table 8.1 The evolution of 2" fast linear focussed PMTs over the past 40 years. The transit-time dispersion (fwhm) is given in the column headed ϵ_{fwhm} , and the field strength at the cathode is in volts per millimetre, under full cathode illumination. The operating voltages are those recommended by the manufacturer. Electronics schematics for measuring ϵ_{fwhm} are given in section 8.6.

Type, manufacturer	$\epsilon_{\text{fwhm}} (\text{ns})$	$dV/ds (\text{V/mm})$
8575, RCA ^a	2.7	8
8850, RCA ^b	0.45	3
56AVP, Philips ^c	1.5	10
9814B, EMI ^d	1.9	3
XP2020, Philips ^e	0.71	15
XP2040, Philips ^f	0.64	15
XP2020UR, Philips ^g	0.47	15
XP2020UR-M, Philips ^h	0.52	15
R1828-01, HPK ⁱ	0.55	50
R2083, HPK ^j	0.37	70

^a The industry standard for fast timing in the 1960s with BeO dynodes, operated with $V_{k-d_1} = 150$ V (Kerns 1967).

^b The construction is the same as for 8575, except for the use of a high-gain first dynode with $\delta_1 = 20$ at 600 V (Hussain and Butt 1980); the 8575 design is the same as the design for C3100E/F, for which Lakes and Poultney (1970) quote the jitter as 0.70 ns.

^c BeO dynodes (Poultney 1972).

^d Electron-optics similar to 8850, with high d_1 gain operated at $V_{k-d_1} = 300$ V (data from Electron Tubes Ltd. 1996).

^e BeO dynodes with an accelerator electrode at d_7 potential; the XP2020 became the standard for fast timing in the 1970s; measurements by Moszyński (1992).

^f Original electron-optics design, as for XP2020, but with a high-gain first stage (Moszyński 1992).

^g Three adjustable timing electrodes, and an additional accelerator, which is adjustable between the d_4 and the d_5 potentials. This PMT includes a first dynode of high-gain.

^h Same as XP2020UR but with an anode screening grid (Moszyński 1993b).

ⁱ High d_1 gain; 12-stage PMT, with the accelerator connected to d_7 (Hamamatsu Photonics K K 2002).

^j High d_1 gain 8-stage PMT, with the accelerator set between d_6 and d_7 potentials. Incorporates coaxial anode output (SMA; Hamamatsu Photonics K K 2002).

8
fidelity. The importance of the extraction field strength to timing can be judged from Fig. 8.4. For those PMTs that follow an s^2 field dependence, there is little acceleration over the first few millimetres of path length, and the relative time spent in this region is considerable, leading to excessive dispersion. In proximity focussed devices with a linear field gradient, the electron velocity is rapid, thus minimizing jitter.

Table 8.2 Other noteworthy fast PMTs.

Type, manufacturer	ϵ_{fwhm} (ns)	dV/ds (V/mm)
R7400, HPK ^{a,g}	0.230	150
R5505, HPK ^{b,g}	0.350	100
R4998, HPK ^{c,g}	0.250	100
PMT413LJ, Photek ^{d,g}	0.015	1200
SiPM, FBK ^e	0.060	—
HPD, HPK ^f	0.440	100

^a The metal package R7400 is the smallest and arguably the fastest of the discrete dynode types, with a diameter of 16 mm and a seated height of 10.3 mm.

^b The first in a range of fine-mesh, proximity focussed PMTs for use in high magnetic fields; diameters range from 25 to 64 mm.

^c 25 mm, 10-stage PMT with accelerator at d_7 potential. The coaxial anode output reduces ringing, compared with conventional designs.

^d Four-stage MCPMT with an active diameter of 13 mm, and a pore size of 10 μm . Ultra-fast PMTs (Photek 1994).

^e Photocathode diameter, 30 μm (Collazuol 2007).

^f Thirteen-inch diameter, hemispherical PMT (Kawai 2007).

^g Timing information taken from data sheets.

8.2.2 Contribution to time dispersion from dynodes

The focus of this chapter has so far concentrated on the front end of a PMT, but recognition must also be given to dynode contributions. Of the conventional multipliers, such as Venetian blind, box-and-grid, circular, and linear focussed types, only the last two will be examined in any detail. The importance of high field gradients at the point of electron emission has been demonstrated in 8.2.1. Venetian blind and box-and-grid dynodes do not have this attribute, thus excluding their use in PMTs intended for fast timing. The electron-optics of the circular and the linear focussed multipliers is essentially the same: a linear focussed set of dynodes may be regarded as an unwrapped version of the circular set.

The challenge faced by the early designers of PMTs in adopting linear focussed dynodes was one of coupling the axial symmetry of the front end to the planar symmetry of the multiplier (Poultney 1972). It should be noticed that the fourth dynode through to the penultimate one have a common profile and orientation with respect to the axis of the PMT, whereas d_1 is angled at about 45° to facilitate efficient photoelectron collection over a finite spot size. Dynodes d_2 and d_3 also have unique shapes and orientation, each designed to collect efficiently and isochronously from their precursors. This asymmetry in the shape and disposition of the first three dynodes of the multiplier is a common feature in the products of all manufacturers. Transit-time dispersion is a statistical as well as an electron-optical

phenomenon, and, as for gain dispersion, the early stages of the multiplication process are crucial. Inter-dynode voltage distributions are particularly non-uniform, with higher field gradients near the trailing edge of each dynode, as illustrated in Fig. 8.11 for two linear focussed PMTs. Photoelectrons land preferentially on this part of the first dynode, by design, and this applies to all subsequent dynodes, except for the ultimate, which is planar. Figure 8.11(a) draws attention to an effect frequently unappreciated; that is, the extraction field at d_1 is determined primarily by the d_3 potential and not by d_2 , as might be expected. This applies to all common dynodes: the magnitude of the extraction field at d_1 is controlled by the presence of

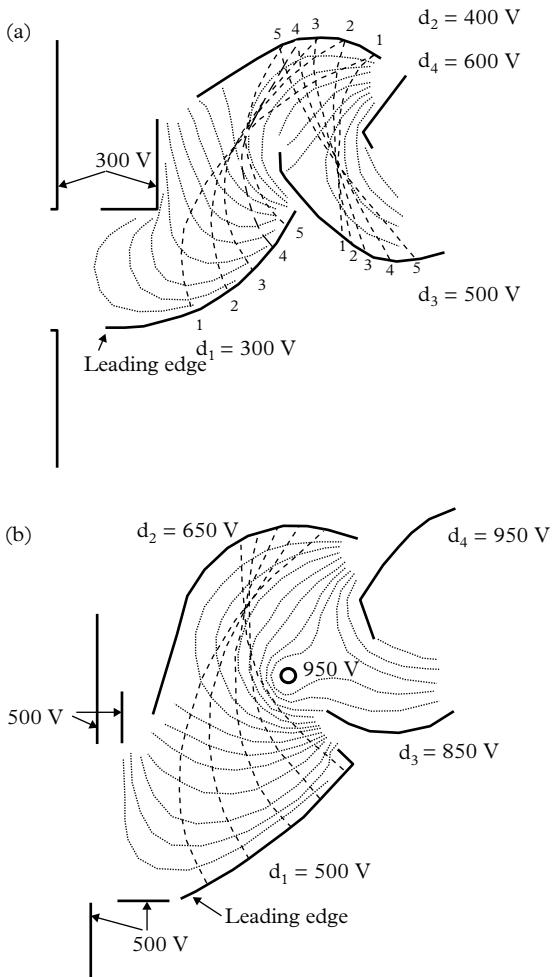


Fig. 8.11. The equipotentials in (a) a Philips XP1020, the predecessor of the XP2020, and (b) the 56AVP. Electron trajectories are depicted by long dashes.

d_{n+2} . In some designs by RCA and Philips, the extraction field on d_1 is determined primarily by the potential on a thin wire parallel to the long dimension of each dynode and located between d_3 and d_2 —this is shown for the 56 AVP in Fig. 8.11 (b). Timing performance can be improved by fine tuning V_{d1-d2} and V_{d2-d3} ; this is achieved by connecting a variable resistor between d_1 and d_3 , with the centre tap connected to the d_2 potential. The resistor value is typically $2R$, where R is value of the common resistor in the voltage divider.

There is a crossover in the space between d_1 and d_2 of the electron trajectories represented in Fig. 8.11(a), and this effect persists between all subsequent dynode pairs. Note that Trajectory 1 has a relatively long transit time to d_2 , due to low initial acceleration, compared to Trajectory 5, for which the extraction field is some ten times stronger. The situation is reversed for trajectories between d_2 and d_3 : for example, Trajectory 1 starts in the high field region of d_2 and lands in the low field region of d_3 . This compensation applies to all pairs of dynodes from the first to the penultimate. There is also a focussing effect produced by the dynode geometry that concentrates the cross section of the electron cloud as it progresses down the multiplier (hence the description ‘linear focussed’). There is further focussing towards the centre in those dynode variants that include side cheeks (see Fig. 5.29(b)).

Electron-optics modelling indicates an interdynode transit time of 2.5 ns and a dispersion of 0.3 ns for the Electron Tubes 9814B. Numerical values of the same order as these follow from (8.7) and (8.9) by taking $s_0 = 10$ mm, and $V = 200$ V, corresponding to an interdynode voltage of 100 V. The reason for assuming 200 V follows from the explanation already given: the field at the n^{th} dynode is primarily determined by dynode number ($n + 2$). The use of (8.7) and (8.9) provides an indication only, since they were derived for a linear field. A reasonable assessment can also be made simply by dividing the transit time by $(n + 1)$, where n is the total number of dynodes.

8.2.3 Transit-time effects at the PMT back end

The last dynode in PMTs is customarily a flat plate, with the anode, in the form of a fine mesh, disposed between it and the penultimate one but located only about 1 mm from it. This construction is known as a reflex anode, and electrons from the penultimate dynode must pass through this mesh to reach the last dynode. The transit-time compensation effect previously mentioned does not apply, although field gradients are high in this region because of compact geometry. Also, usual practice when dealing with pulses exceeding a peak of 10 mA is to apply a tapered voltage distribution at the back end, further enhancing potential gradients. It is known that timing performance may be improved by taking the signal from the last or penultimate dynode rather than from the anode (Bengtson and Moszyński 1982; de Vries and Kelling 1988). To understand the reason for this, we must consider the contribution from induced anode current; this is generated by the motion of the pulse of electrons in transit between the last two dynodes.

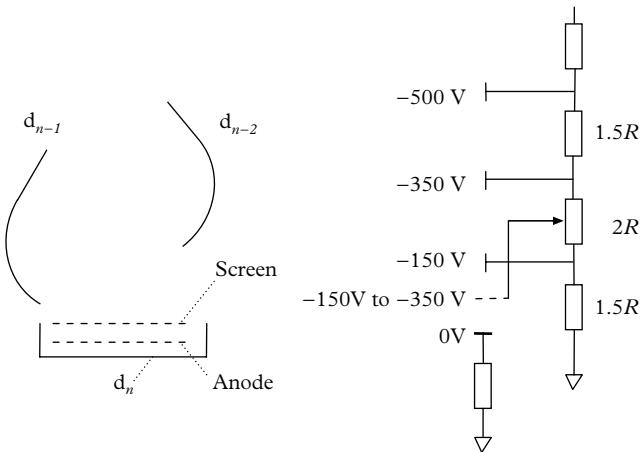


Fig. 8.12. The screen grid introduced by Philips in the XP20Y/DA fast PMT, showing the biasing arrangement. The action of the screen is twofold. First, it intercepts the induced current, which would otherwise flow to the anode; this current originates from secondary electrons moving between d_{n-1} and the screen. Second, the screen grid hastens the collection of electrons from d_n .

8

The current induced by an electron moving normally at velocity v , at a distance d from the anode mesh is $i_{in} = ev/d$. This refers to a uniform field and follows from the work by Spangenberg (1948). The signal is bipolar, with positive current flowing from the anode to ground as the electrons approach it, and then reversing, once the pulse has passed through the anode mesh on its passage to d_n . According to (8.8), the direct electron signal from the last dynode follows the induced one after an interval of about 0.5 ns, with a jitter of about 0.1 ns.

Philips introduced a fine screening mesh between d_{n-1} and the anode, biased as shown in Fig. 8.12. Secondary electrons moving between d_{n-1} and the screen are prevented from inducing current into the anode because of the shielding action of this mesh; it is only once electrons have passed through the screen that current is induced. Transit times are short, and the parasitic signals are fast. Biasing the screen hastens collection by turning back the electrons yet to be collected by the anode. The amplitudes of the induced and the direct signals are different and not clearly delineated, even in the fastest PMTs. However, there are visual signs, in some PMTs, of the induced component at the very start of the anode signal. This precursor, if of sufficient magnitude, can obviously cause early triggering, generating added dispersion. An improvement in timing performance by up to 20 % has been reported by Moszyński *et al.* (2004) in a modified XP2020, the XP20Y0/DA, which includes a screen electrode.

8.2.4 Ultra-fast PMTs

A new generation of high-energy physics accelerators (the Large Electron Positron Collider) came into operation in the 1980s. Some of the detector requirements were met with vacuum photodiodes and phototriodes, which were able to perform satisfactorily in a high-magnetic-field environment. The low gain of these devices restricted their use to detectors in which light levels were high and, to meet other, lower-light-level demands, Hamamatsu introduced a fine-mesh range of PMTs with up to 19 stages of gain (R6054). These are proximity focussed devices, with short interelectrode spacing, and timing performance is consequently comparable with that of the best linear focussed devices. The 10 mm metal-can PMTs in the R7400 range are among the smallest detectors available. They have the same attributes as the fine-mesh variety, regarding close electrode spacing and fast timing.

8.3 PMT output signal

8.3.1 Signature of the anode signal

The shape of the anode signal has a bearing on timing and resolution of a PMT. To understand this, we need the time signature for single-photon excitation. The response to a multi-photoelectron event of any given form may be derived by superposition. In practice, we are restricted to the use of functions that are readily integrated, as will become obvious. The following characteristic pulse shape, $i(t)$, is assumed at the anode and prescribed for every current pulse initiated by a single photoelectron:

$$i(t) = -(q/\tau_p) \exp(-t/\tau_p), \quad (8.14)$$

with $q = g e$, where g is the gain of the multiplier, e is the electronic charge, and τ_p the time constant of the trailing edge of the anode pulse. Representing the trailing edge by an exponential decay, and the leading edge by a step, is in accord with observation. The light output from the majority of commercially available scintillators is adequately described in a similar way by an exponential decay of single time constant, τ_s . Some scintillators exhibit a slow component but of sufficiently low amplitude to be ignored for present purposes. If N is the total number of photoelectrons in an event,

$$f'(t) = (N/\tau_s) \exp(-t/\tau_s), \quad (8.15)$$

where $f'(t)$ is the rate of production of photoelectrons at time t , and the number of photoelectrons produced between t and $t + dt$ is $f'(t)dt$. These make the following contribution to the output at time t' :

$$dI(t') = f'(t)dt \times i(t' - t),$$

where, referring to Fig. 8.13, t' is always $\geq t$.

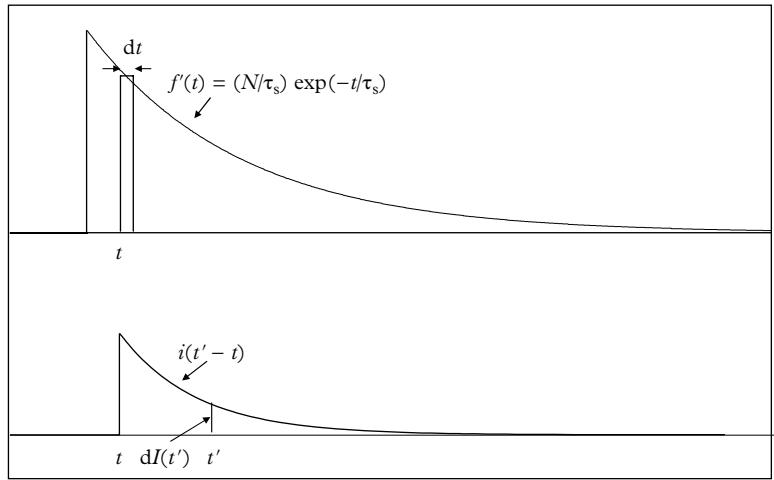


Fig. 8.13. In the signal-forming process at the anode, each photoelectron produced in the time interval t to $t + dt$ contributes $dI(t')$ to the output at time t' ; $I(t')$ is the sum of all such contributions from decays falling within $0 < t < t'$ and is given by (8.17). The anode signal is negative-going, but is shown positive, for illustrative convenience. Substitution of λ_s for $1/\tau_s$ is also made for this purpose.

8

Integrating all contributions from decays initiated between 0 and t' , we have

$$\begin{aligned}
 I(t') &= - \int_0^{t'} N \lambda_s \exp(-\lambda_s t) q \lambda_p \exp(-\lambda_p(t' - t)) dt \\
 &= -N q \lambda_s \lambda_p \exp(-\lambda_p t') \int_0^{t'} \exp(-(\lambda_s - \lambda_p)t) dt \\
 &= \frac{-Nq}{(\tau_s - \tau_p)} \left(\exp(-t'/\tau_s) - \exp(-t'/\tau_p) \right), \\
 I(t) &= \frac{-Q_0}{(\tau_s - \tau_p)} \left(\exp(-t/\tau_s) - \exp(-t/\tau_p) \right).
 \end{aligned} \tag{8.16}$$

The dummy variable t' has been replaced by t in (8.16).

As $V_0(t) = R I(t)$, and $\tau_p = RC$,

$$\begin{aligned}
 V_0(t) &= \frac{-RQ_0}{(\tau_s - \tau_p)} \left(\exp(-t/\tau_s) - \exp(-t/\tau_p) \right) \\
 &= \frac{-Q_0 \tau_p}{C(\tau_s - \tau_p)} \left(\exp(-t/\tau_s) - \exp(-t/\tau_p) \right).
 \end{aligned} \tag{8.17}$$

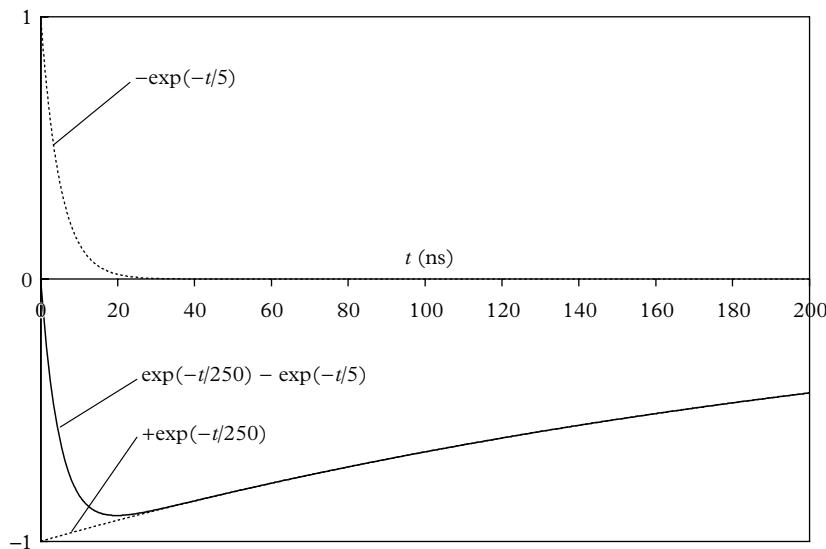


Fig. 8.14. The effect of two time constants, of different sign, on the leading edge of a PMT output pulse, derived from (8.16), with $\tau_s = 250$ ns, and $\tau_p = 5$ ns.

Equation (8.17) is identical to (14.4) (derived by a different approach), as it must be. The derivation of (8.17) is known mathematically as folding, a process already familiar in the statistical treatment of 4.4. The purpose of the presentation given here is to highlight the way in which the final signal is synthesized from individual photoelectron events. We note that the output pulse comprises two exponential functions: one representing the scintillator, τ_s , and the other the PMT signature, τ_p . The effect of this is to produce an anode signal with a leading edge of finite rise time, as depicted in Fig. 8.14. The importance to timing is obvious and, as a broad generalization, the faster the leading edge, the better is the timing.

The waveforms in Fig. 8.14 take no account of the quantized nature of charge. In reality, there is a statistical overlay on the smooth profiles, and this becomes more pronounced for signals that originate from just a few photoelectrons. In 8.3.2 we simulate the statistical nature of scintillator light, also taking account of the fluctuations in pulse height from one anode signal to the next. For this purpose, we re-examine the example represented in Fig. 8.14.

8.3.2 Simulation of photoelectron time intervals

The light output from scintillators may be described by the exponential decay of a single time constant, τ_s . This applies to the majority of commercially available scintillators, with a minority better described in terms of two decay time constants known as the ‘fast’ and the ‘slow’ components. Although the present discussion is

restricted to just one time constant, τ_s , we can assume that the detection of these photons, as photoelectrons, follows the same law:

$$f'(t) = \frac{dQ}{dt}(t) = (N/\tau_s) \exp(-t/\tau_s) \quad (8.18)$$

with Q and N scaled by the photocathode QE. By restricting the discussion to small-area and small-volume scintillators, it follows that the emission of a photon and its detection at the photocathode may be assumed isochronous. This is satisfied in practice, providing the scintillator dimensions are restricted to a few centimetres only. Integrating (8.18) from $t=0$ to ∞ gives N , the total number of photoelectrons detected, as required; $Q(t)$, the number of photoelectrons emitted in the time interval 0 to t , follows from the integration of (8.18):

$$f(t) = Q(t) = N \left(1 - \exp(-t/\tau_s) \right). \quad (8.19)$$

The 10 to 90 % rise time is determined from (8.19), arriving at $t_r = \tau \ln(9) = 2.2\tau$. Equations (8.18) and (8.19) describe the average emission, and detection, of light from a scintillator. This is illustrated in Fig. 8.15, where the detection timeline is plotted for a total of $N = 100$ pe, using (8.19). The points follow a smooth curve, and we note that the intervals between detections increase with t , strictly in accordance with (8.19).

The concept of average in describing light emission from scintillators limits our ability to explain actual processes occurring in nature. Light emission from a

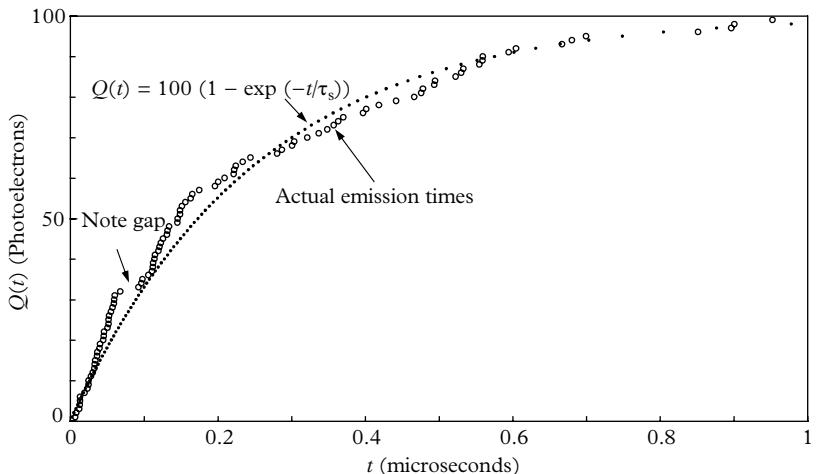


Fig. 8.15. Illustrating the detection of 100 pe from a scintillator with a single decay time constant of $\tau_s = 250$ ns. The small dots describe the path followed in the absence of statistical effects. The intervals between the dots increase with time, in accordance with the exponential distribution. Any particular decay sequence of 100 pe will follow a random path, such as that shown by circles.

scintillator, following excitation by a gamma ray, for example, is a statistical process requiring Monte Carlo methods to simulate real processes. The statistical simulation of (8.19) is shown in Fig. 8.15, where the random walk of the emission times around the path of (8.19) is nicely illustrated. Plots such as that in Fig. 8.15 are obtained by selecting a fixed number of random detection times weighted in accordance with (8.19); they are then ordered in time and plotted as shown. Every repeated simulation is different, and the deviation from the average curve can be substantial, because there is no mechanism for reducing the walk once it has strayed from this path: at any point in the progression, the next point on the curve is as likely to be further from the average line as to be closer to it. This is a well-known effect in a repeated coin-tossing game, known as change of sign and easily simulated using a spreadsheet function. A theoretical treatment is provided by Feller (1957). The appearance of noticeable gaps has a significant effect on the pulse shape of real events (e.g. see 8.3.2, Fig. 8.17).

8.3.3 Simulation of multiplier noise and jitter

A scintillator pulse can be simulated by starting with the time interval distribution shown in Fig. 8.15. A characteristic PMT output pulse, assumed to be of the form (8.14), is generated by each of 100 pe in an emission-time simulation. This set of anode sub-pulses, when summed, constitutes the simulation of a 100 pe scintillation pulse, but on the assumption of a perfect multiplier. The simulation is easily refined by including the statistical effects of noisy gain: in place of taking a fixed

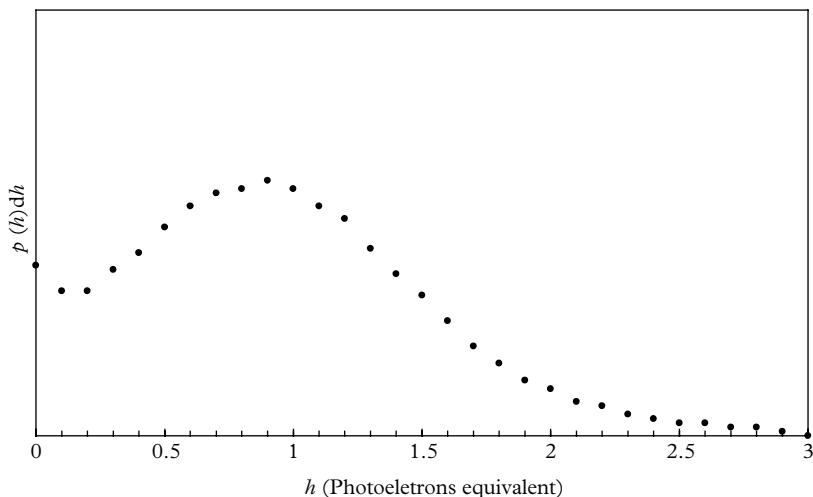


Fig. 8.16. An SER distribution measured at the anode of a PMT. The mean of the distribution is unity but output signals range in size from a fraction of a photoelectron equivalent to about 3 photoelectrons equivalent.

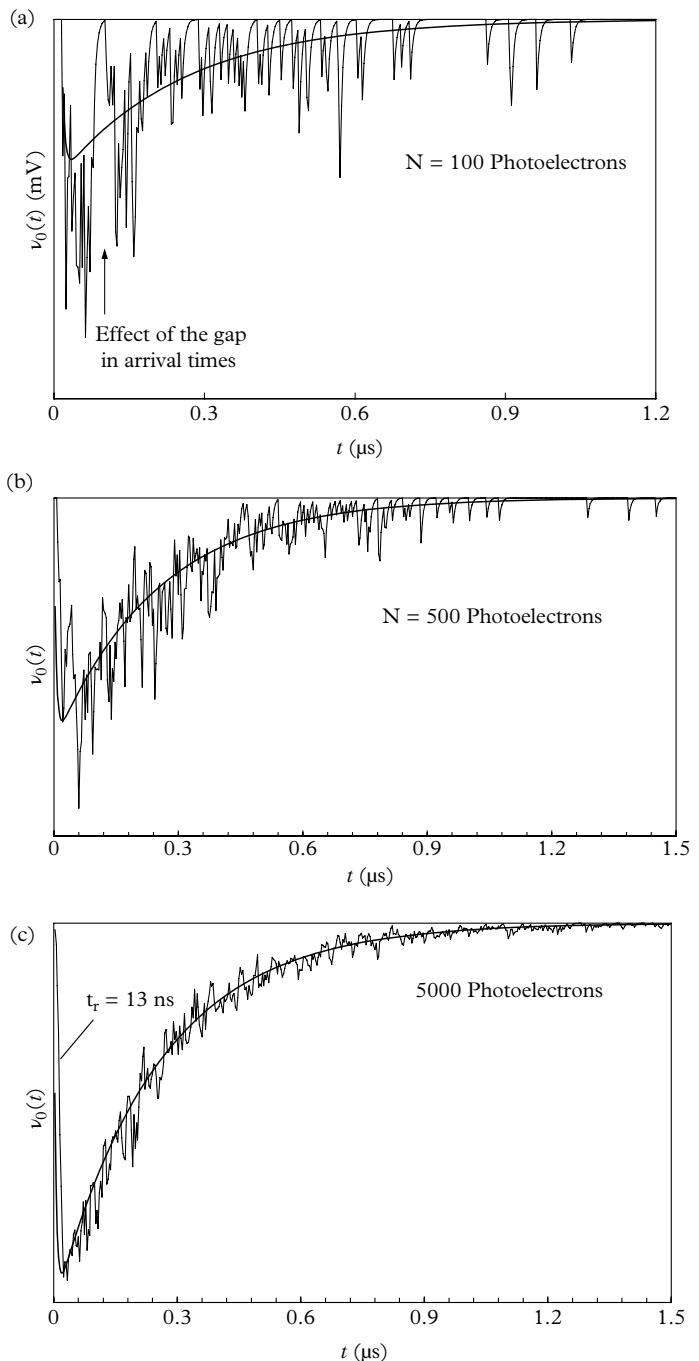


Fig. 8.17 Examples of synthesized PMT pulses of 100, 500, and 5000 photoelectrons generated by NaI(Tl) allowing for statistical fluctuations in detection times of the photons and for fluctuations in the magnitude of the anode output—the SER. The photoelectron creation time for (a) is taken from Fig. 8.15. The traces in (b) and (c) refer to higher photoelectron numbers, as illustrated. The simulation conditions are $\tau_s = 250$ ns; $\tau_p = 5$ ns; and $\epsilon_{PMT} = 5$ ns.

amplitude of $-q/\tau$ for each constituent pulse, this quantity is allowed to fluctuate from event to event, in accordance with the distribution shown in Fig. 8.16; allowance is made for transit-time jitter, assumed normally distributed, with a standard deviation of $\epsilon_{\text{PMT}} = 5 \text{ ns}$. Although functional forms such as a Poisson or a normal distribution could have been chosen to represent multiplier noise, it is more realistic to select events from a measured SER, such as the one shown in Fig. 8.16, if the intention is to obtain a representative simulation.

The simulation steps outlined above lead to the signal shown in Fig. 8.17(a). The solid line represents (8.16): the signal that would be generated in the absence of statistical effects. The area under both curves is the same, but the characteristic shape of the scintillation pulse is masked by noise. The gap highlighted in Fig. 8.15 has a noticeable effect on the decay profile, which actually touches the baseline after $\sim 0.1 \mu\text{s}$ because of the discontinuity. The group of pulses appearing four time constants after the start of the pulse is noteworthy.

The simulation in Fig. 8.17(a) is typical of a 10 keV NaI(Tl) pulse with further examples for 500 and 5000 photoelectrons equivalent in Fig. 8.17(b) and (c). These are less noisy than the first simulation, simply because of the number of photoelectrons assumed, but the imprint of single photoelectrons is still evident, particularly in the tails. The smooth curves shown follow from (8.17). There are late photon arrivals, even six time constants after the start, making retrigerring on the same event a possibility. The rise time of the smooth pulse is 13 ns, calculated on the basis of 10 to 90 % of pulse height. It is clear by inspection of the traces of Fig. 8.17 that this particular timing parameter is easily corrupted by noise on the leading edge. The 500 pe signal in Fig. 8.17(b) illustrates this effect, whereas the less noisy trace of Fig. 8.17(c) is relatively unaffected by statistical fluctuations on its leading edge.

Monte Carlo simulations given in Fig. 8.17 show the effect of quantization, as noise in signal superimposed on the output trace. Individual photoelectron contributions occurring at a time equal to five time constants after the start of the signal are evident in the tail of all simulations. These late arrivals can be particularly troublesome when counting scintillation pulses and under certain circumstances (such as insufficient dead time) will lead to double counting. The degree of noise corruption revealed by the simulations is perhaps surprising, especially concerning pulses of low photoelectron number. Examples of actual pulses generated by a NaI(Tl) scintillator are investigated in the next section to compare with the simulations in Fig. 8.17.

8.3.4 Oscilloscope traces for NaI(Tl) signals

The level of noise on the profiles given in Fig. 8.17 may appear excessive and contrary to expectation. This is because the observed signals are often smoothed, either by the time constant of an anode load, or by the action of preamplifiers and main amplifiers. That the actual anode signal indeed resembles those illustrated is easily verified by comparison with recorded gamma ray interactions in a NaI(Tl)

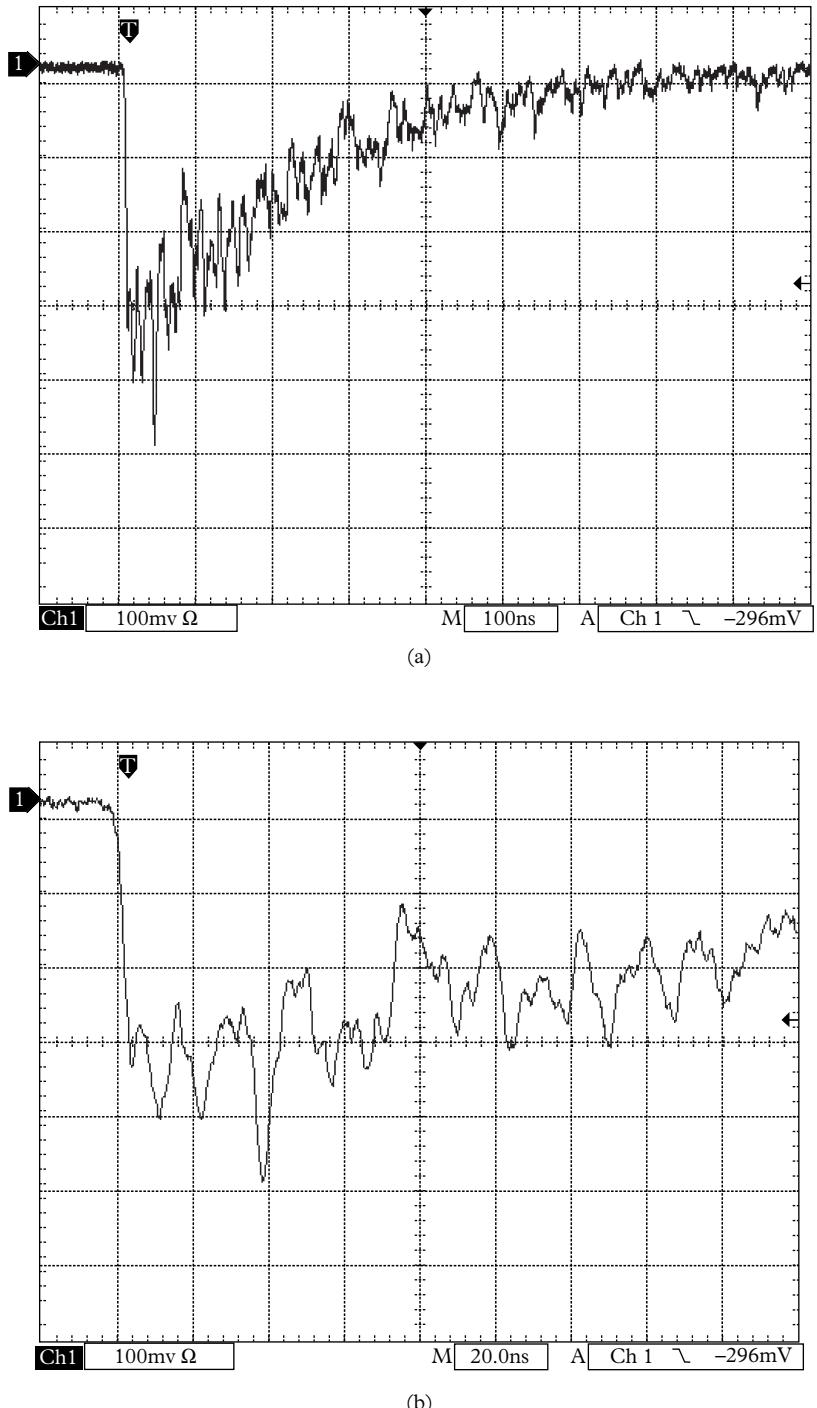
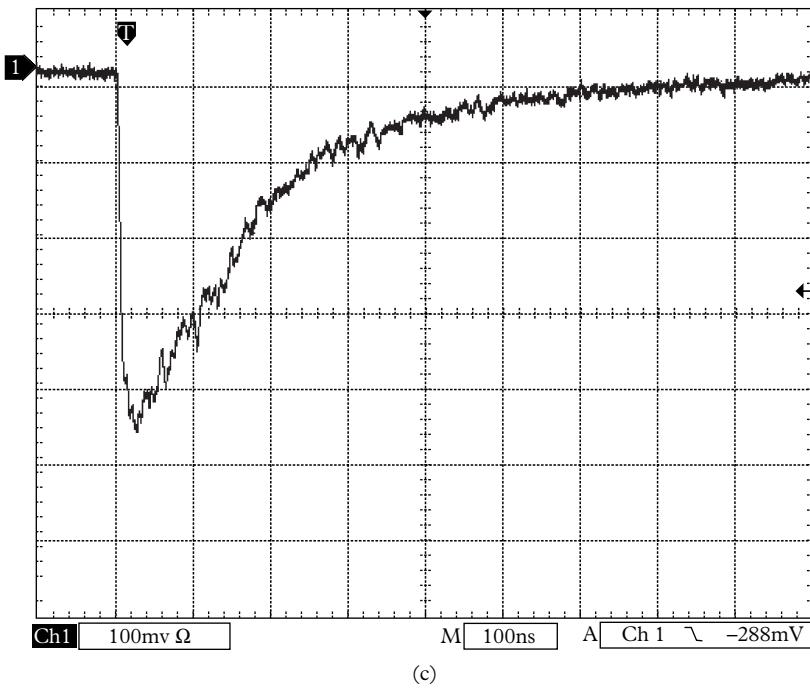
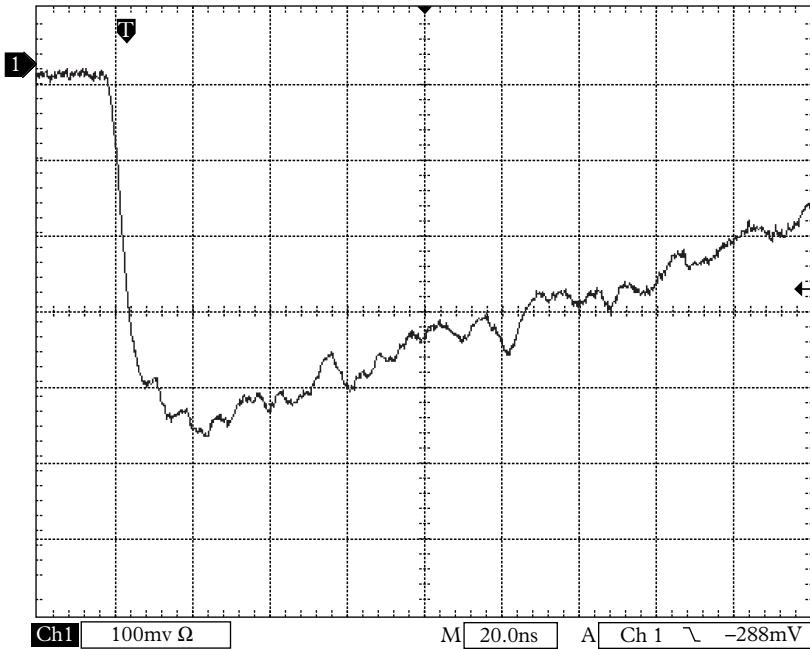


Fig. 8.18. Traces of NaI(Tl) pulses using a PMT of rise time 1.5 ns, recorded with a fast oscilloscope. (a) An event of about 1000 pe exhibiting a leading edge of less than 20 ns. (b) An expanded version of (a), showing single-electron pulses with an amplitude of ~ 100 mV, and an fwhm of ~ 4 ns. Traces (c) and (d) show a 5000 pe event for the same oscilloscope settings, while the PMT gain has been reduced by a factor of 5. Note that the leading edge rise time of about 10 ns corresponds to a time constant of ~ 5 ns, based on $t_r = 2.2\tau$, for an exponential function.



(c)



(d)

Fig. 8.18. Continued

scintillation crystal. A fast PMT operating into a $50\ \Omega$ matched load is necessary for this purpose. The coefficient in (8.17) can be taken as $\sim 10\ \text{pe/keV}$ and, on this basis, gamma rays from ^{137}Cs and ^{57}Co sources will produce signals comparable in magnitude with those of the simulations in Fig. 8.17(b) and (c). The level of noise exhibited by the simulations is seen to be comparable with those of the oscilloscope traces with single-electron pulses clearly in evidence in Fig. 8.18(b). Equation (8.17) can be verified by superimposing the signal trace of Fig. 8.18(d) on the curve representing (8.17), as in Fig. 8.19.

8.3.5 Scintillator rise and fall time

There is considerable literature on the measurement of scintillator decay time, τ_s , but until recently little on the topic of rise time. Two papers by Derenzo *et al.* (2000) and Weber *et al.* (2000) report on measurements made with a fast MCPPMT. The behaviour of inorganic scintillators fall into two groups: the intrinsic scintillators (undoped), such as CdWO_4 , BGO, and CsI, have a rise time $<30\ \text{ps}$; the activated scintillators, such as $\text{CaF}_2(\text{Eu})$, $\text{CdS}(\text{Te})$, and $\text{YAlO}_3(\text{Ce})$, are slower, but under $240\ \text{ps}$. $\text{LaBr}_3(\text{Ce})$ is quoted by Moszyński *et al.* (2006) as $1\ \text{ns}$. Rise times for $\text{NaI}(\text{Tl})$, $\text{CsI}(\text{Tl})$, and $\text{CsI}(\text{Na})$ are $>10\ \text{ns}$, compared with intrinsic CsI, which has $\tau_r \approx 30\ \text{ps}$ (Derenzo *et al.* 2000). Measured waveforms by the present author indicate a rise time in the region of 5 to 10 ns for $\text{NaI}(\text{Tl})$ (Figs 8.19 and 8.18(d)). Lerche and Phillion (1991) quote a rise time for the plastic scintillator BC-422 of $<20\ \text{ps}$.

Derenzo *et al.* (2000) cover a wide range of scintillators for both rise and fall time. They also demonstrate most convincingly the contribution to rise time from multiple light paths in the scintillators. For example, a 10 mm cube sample of PbWO_4 has a

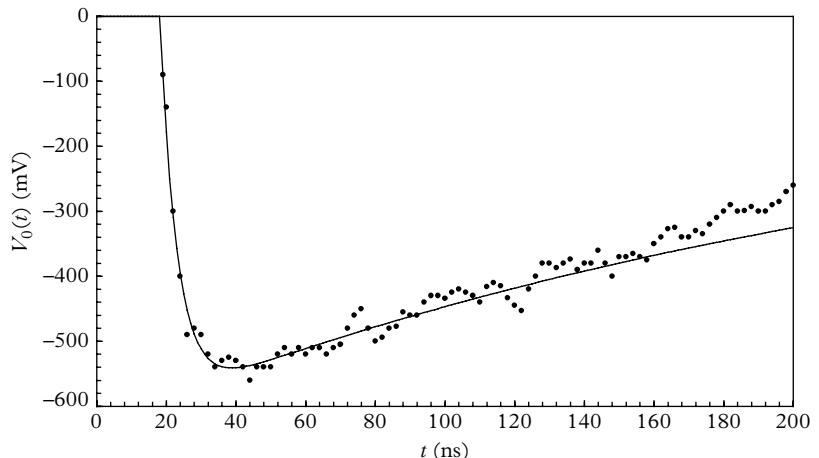


Fig. 8.19. Pulse height readings taken from Fig. 8.18(d), superimposed on the curve representing (8.17) with parameter values $\tau_s = 250\ \text{ns}$, and $\tau_p = 5\ \text{ns}$.

rise time of ~ 500 ps when clad with white diffuser which reduces to < 100 ps when all faces, except the one coupled to the detector, are blackened. Another example of the practical consequences of finite rise time has been reported by Rosza *et al.* (2009) in a technical note, showing that time dispersion increases, essentially uniformly, from 110 ps to 260 ps for $\text{LaBr}_3(\text{Ce})$ crystals ranging in length from 5 to 51 mm (in all crystals examined, the length and diameter are the same).

8.4 Timing

The requirement is one of determining the arrival time of a signal, usually at the anode, but sometimes it is picked off at a dynode close to the anode. Ideally, this should be done without adding to the time dispersion already contained in the signal. An electronic circuit such as a TTL, emitter-coupled logic (ECL), or low-voltage differential signalling (LVDS) circuit, for example, is required to produce a fixed-amplitude and fast-rise-time signal for each event exceeding a set threshold. Such circuits are known as fast-timing discriminators or time pick-off circuits; in these circuits, both timing and discrimination are performed in the same electronics module. There are many considerations in choosing the optimal threshold setting for discrimination, as will be discussed in 8.6. In practical terms this translates into setting a threshold for output signal pulses as low as is practical, for both single- and multi-photoelectron signals. To give an indication of the sensitivity required, it is common practice to operate PMTs in timing circuits at a gain within the range 10^6 to 10^7 . Assuming the latter figure, then 1 pe will generate an anode pulse of charge 1.6 pC. A typical fast PMT produces a pulse of width ~ 2 ns for single-photon excitation, leading to a peak amplitude of

$$\sim \frac{1.6 \times 10^{-12}}{2 \times 10^{-9}} \times 50 \text{ V},$$

for a 50Ω load. The peak pulse height is accordingly of the order of 40 mV. This is barely sufficient to trigger a standard, commercially available, NIM timing discriminator, for which the minimum threshold is typically 50 mV. These rough calculations point to the need for a fast $\times 10$ amplifier to precede the timing discriminator, increased PMT gain, or a combination of both.

8.4.1 Single-electron time response

The shape of the anode signal influences timing performance and in the preceding sections the mathematical convenience of an exponential function was assumed. For the considerations that follow, we need to adopt a form that includes both finite rise time and fall time. A PMT output signal, $i(t)$, whatever its analytical form, can be characterized in terms of the PMTs response to a single photoelectron. This type of

input satisfies the definition of a δ function and it then follows, from the well-known theorem in network analysis, that the output response of the PMT to any multi-photoelectron function can be determined from $i(t)$. There are two types of distribution to consider: the charge response and the time response, both of which are statistical in nature. An SER is the pdf of the output charge distribution of single-photoelectron initiated events, and it is usually referred to as a pulse height distribution; it relates only to output pulse size, and the shape of the pulse is immaterial in this case. A single-electron time response (SETR) describes the distribution in arrival times of the electron cascade that constitutes $i(t)$, without reference to the pulse height distribution; the spread in arrival times, at the anode, is quantified by λ , which is related to the width of the output pulse. The SETR is therefore the time signature of a single-photoelectron event measured at the anode, or at a dynode if the signal is picked off at that location. It is important to be clear on the difference between the SER and the SETR, because it is not unusual in the literature to find the term SER used to describe both these distributions, and sometimes an intelligent guess has to be made as to which applies.

The signal amplification process starts with a δ -function input to the first dynode, from which δ_1 secondary electrons are generated. This set of electrons is released from d_1 without time dispersion but, because of the spread in initial energies and launch angles, electrons arrive at the second dynode at different times. If we assume that the arrival times at d_2 follow an exponential distribution, it follows that the arrival time distribution at d_3 will be an exponential folded with itself, and so on for the remaining dynodes. The solution to this problem is the same as that for secondary emission yield for a set of dynodes, all of which are also assumed to follow an exponential distribution (see (4.51)). The spread in arrival times at the n th dynode, $i_n(t)$, is thus

$$i_n(t) = t^n \exp(-at), \quad (8.20)$$

as depicted in Fig. 8.20. It will be noted that (8.20), once plotted, does indeed bear the essential characteristic shape of a PMT output signal: the leading edge being always faster than the trailing edge, with an fwhm increasing with n . It will be noticed that the distribution becomes more symmetrical with increasing n , and in the limit it becomes normal.

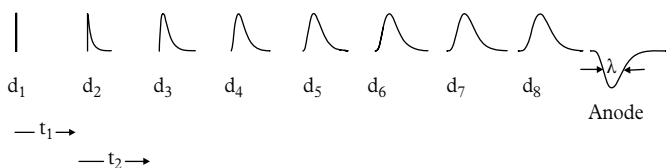


Fig. 8.20. The evolution of the signal pulse shape as it progresses down the multiplier (the increase in amplitude with n , the number of dynodes, has been removed by normalization). The width of the signal, λ , gives an indication of expected timing performance, known as jitter.

8.4.2 Transit time, T

The question of measuring time dispersion in T should be investigated in a conceptual way prior to any discussion on practical circuits. The transit time is the difference in time between the creation of a photoelectron and the arrival of the resulting anode signal. Transit time is a statistical quantity and we refer to the average time taken as T . It manifests itself in two ways: first, as a delayed signal, which is not normally of particular interest or inconvenience in PMT applications; and, second, as a source of timing uncertainty. In multi-detector systems, it may be necessary to equalize the delays pertinent to each detector to synchronize their output pulses; this is easily done through the use of a few metres of coaxial cable. Alternatively, stepped delay units are commercially available and offer the convenience of precise delays from 1 to 32 ns, in steps of 1 ns. The other manifestation of T , of paramount concern, is the dispersion in arrival time, from one signal pulse to the next. This is illustrated in Fig. 8.21, showing a sample of three events. All assume identical shape (same λ) but arrive at different times relative to a common 'start' pulse. Because of the statistical nature of gain, every output pulse has a different shape and hence different λ . Clearly, T is affected by λ and particularly in its variation from pulse to pulse, but the exact relationship is obscure and complicated.

Repeated measurement leads to the generation of a histogram for the transit-time dispersion in T . It is not unusual to accumulate in excess of 10,000 events, with the results displayed as a histogram (Fig. 8.22); T is simply the mean of T_1, T_2, \dots, T_n . For these parameters to have meaning, some point on the output pulse must be taken as reference, or 'mark', as it is also called. There is a choice, in

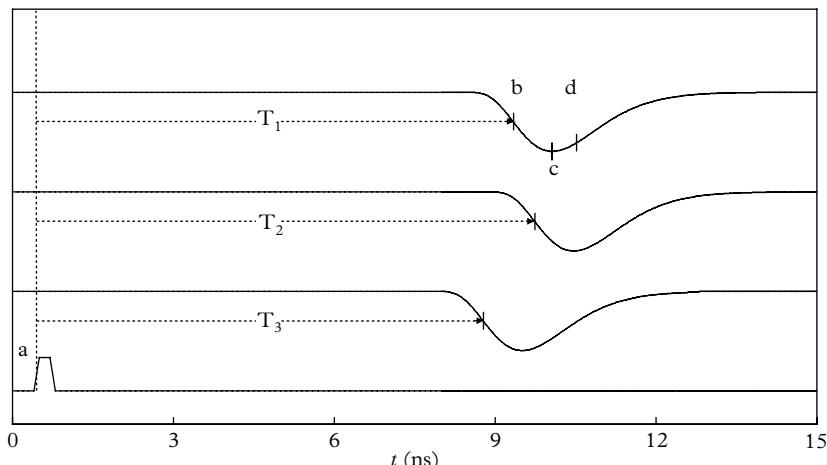


Fig. 8.21. The concept of transit time, and three of the ways, (indicated by a, b, and c), in which it can be specified.

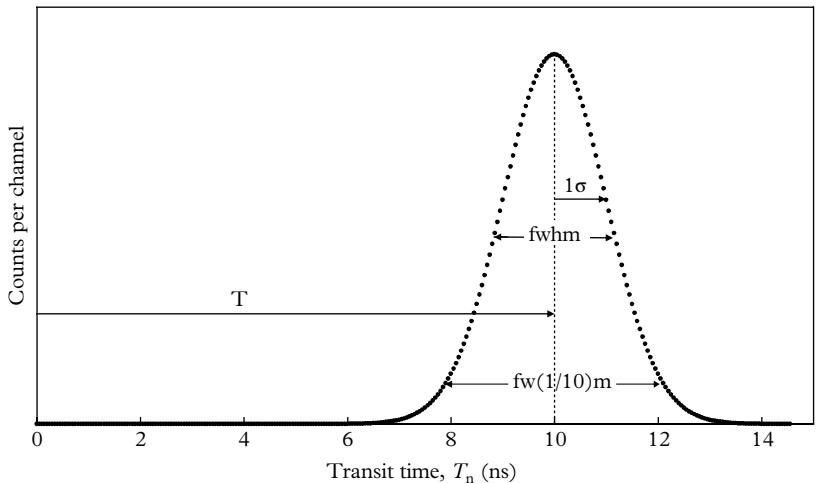


Fig. 8.22. An idealized distribution of timing events illustrating the ways in which the jitter, ε_{PMT} , may be presented. Manufacturers and users are not always clear as to which parameter applies: ε_{PMT} being smaller than the fwhm, is sometimes preferred by manufacturers; ε_{PMT} refers to one standard deviation, σ . It is customary to use a log scale for the ordinate.

8

Fig. 8.21: the centroid (d); the peak (c); or the 50 % rise-time point (b). In this example, the single-electron pulses have a common profile, and the choice of reference point appears unimportant, which is not so in practice.

The standard deviation of the dispersion in T is usually referred to as σ or ε_{PMT} . It may be expressed as a variance, which is the square of these quantities. Alternatively, the fwhm of the distribution may be used. The link between these two statistical quantities is the obvious one and, for a normal distribution, the fwhm and the standard deviation are directly related to one another through

$$fwhm = 2\varepsilon_{PMT}(2\ln 2)^{1/2} = 2.35\varepsilon_{PMT}. \quad (8.21)$$

Because of the ease of calculation, this relationship is frequently used, even when the experimental distribution is noticeably asymmetrical in shape: only a ruler is required to determine the fwhm of a distribution, whereas determining ε_{PMT} requires computation. Sometimes the full width at one-tenth maximum (fw(1/10)m) is quoted as a measure of consistency in a normal distribution.

In many PMT applications, the concern is the timing performance obtained for a specific light source. With regard to PET scanners, for example, the time dispersion observed with light pulses from a particular scintillator which is excited by coincident annihilation radiation (0.511 MeV) is the parameter of interest. However, in order to compare the performance of PMTs and as a basis for selection, the time dispersion for single photoelectrons is universally accepted as

the appropriate measure. The multi-photoelectron sources mentioned previously can always be attenuated to generate just a single photoelectron, in the manner of 4.1.1. By adjusting the attenuation in a known way, one can produce a time dispersion plot showing this parameter as a function of the number of photoelectrons per pulse, if required.

Consider single-photoelectron events that produce an identical pulse for every photoelectron. Specifically, the assumption is that the shapes of all pulses regarding rise time, fall time, and total charge are identical from pulse to pulse. Because the pulses are identical, the assignment of time dispersion, with reference to some average value for the transit time, will be independent of the timing reference point chosen—any one of b, c, or d in Fig. 8.21. Clearly, the average transit time does depend upon the choice of reference point, but this is unimportant since we are only concerned with the deviation from the average.

A pulse amplitude discriminator is an electronic circuit that gives a fast digital output for all input signals that exceed a set threshold. Given such an ideal threshold discriminator, setting b close to the baseline but sufficiently above the noise would seem a sensible choice. However, real PMTs excited by single photons produce pulses with a range of sizes extending over a dynamic range of about 10:1 (see e.g. Fig. 8.16). The situation depicted in Fig. 8.21 is ideal but clearly unrealistic, and we must consider the contribution that pulse amplitude variations have on the arrival time of any single-photon-initiated event. For the discussion that follows, we assume the signal profile $i(t) = t^n \exp(-at)$, with $n=10$, together with a range in pulse heights. Key features such as a finite rise time and an extended tail (see Fig. 8.23) are embodied in this function. It is immediately

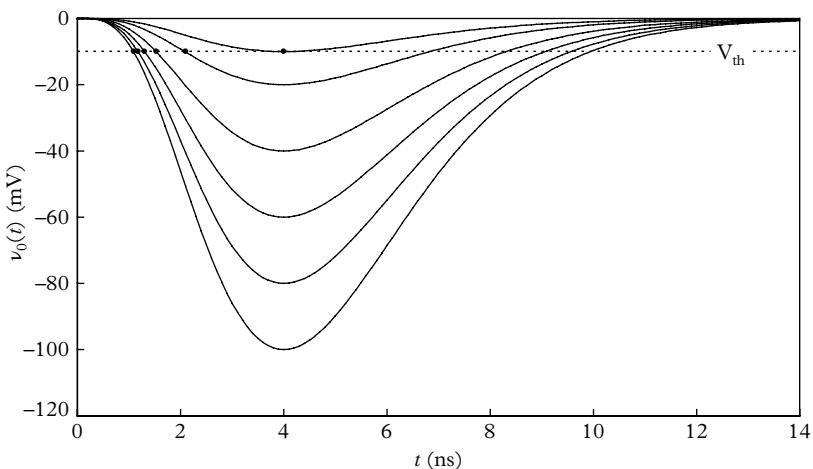


Fig. 8.23. Illustrating the phenomenon of time walk with threshold discriminators. The six pulses have a common shape but different amplitudes, thus crossing the set threshold, V_{th} , at different times, indicated by the solid circles.

obvious that larger amplitude pulses appear to arrive earlier than smaller ones if the arrival time is taken as the instant at which each pulse crosses a fixed threshold, V_{th} . This phenomenon is known as ‘time walk’ or ‘time slewing’, and the way to reduce the effect in this type of discriminator is to lower the threshold level, which incurs another problem—false triggering on noise. The extent of the walk will be of the same order as the rise time of the PMT pulse, that is, in the range of about 1 to 10 ns. The time at which each waveform crosses the threshold V_{th} is indicated by the small dots in Fig. 8.23. Observing that the time at which each pulse reaches its peak is amplitude invariant, it is obvious that, if a method based on this feature could be devised, detection would be isochronous for all pulse amplitudes. This thought process leads to the technique known as zero crossing, which is based on a differentiated version of the PMT output pulse.

8.4.3 Effect of amplitude fluctuations on jitter

The SER characterizes the fluctuation in total output charge, q , produced for each detected photon (see (8.17)). In the special case of an ideal multiplier obeying Poisson statistics, the relative variance in q is given by $1/(\delta-1)$ (see (4.49)). It is assumed in this expression that all dynodes have the same gain, but a refinement of the formula allows for individual gains, $\delta_1, \delta_2, \dots, \delta_n$. However, it is preferable to continue the discussion on the basis of measured SERs because theoretical ones lead to optimistic conclusions, failing to predict the high proportion of small pulses observed in practice. Assuming single-photon excitation with a standard shape for the PMT output, shown in Fig. 8.24, we can transform the effects of fluctuations in amplitude, represented by $V_1(t)$ and $V_2(t)$, into fluctuations in time.

It follows from geometrical considerations that

$$\Delta V(t) \approx \frac{dV(t)}{dt} \Delta t$$

or

$$\Delta t = \Delta V(t) / \frac{dV(t)}{dt}, \quad (8.22)$$

where $dV(t)/dt$ is the slope of the average signal profile as it crosses the discriminator threshold at $t=T$. Note that ΔV and Δt may be jointly taken as standard deviations or relative variances, or any other measure of dispersion, from which it follows that the relative variance in the transit-time jitter due to the relative variance of the SER is

$$\epsilon_t^2 = \epsilon_A^2 / \left(\frac{dV(t)}{dt} \right)^2 \quad (8.23)$$

at $t=T$.

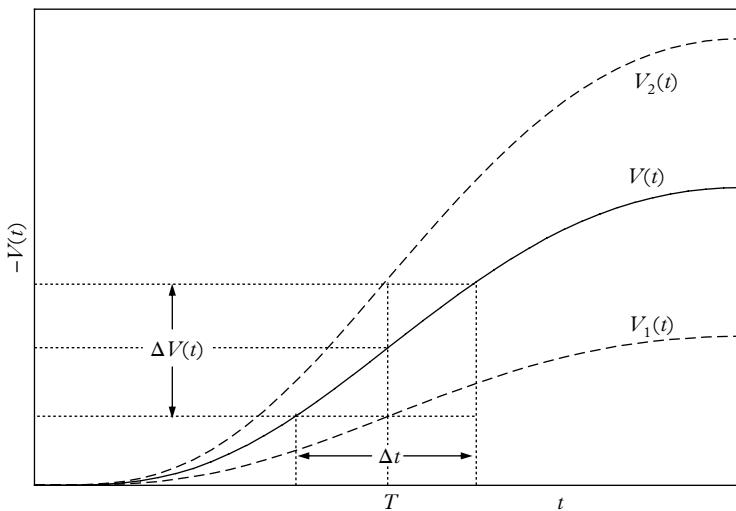


Fig. 8.24. The output from a PMT for single-photon excitation, showing the leading edge only. The solid curve illustrates the relationship between amplitude noise, $\Delta V(t)$, and time noise, Δt .

If the rise time, t_r , is linear, and the fwhm of the SER is Δ , then the rms jitter is

$$\varepsilon_t = \varepsilon_A t_r = \Delta t_r / 2.35. \quad (8.24)$$

8

8.4.4 Effect of charge sensitivity on jitter

Once a signal crosses the threshold level of a discriminator, an additional small quantity of charge, Δq , is necessary to trigger the device. The way in which this leads to a contribution to walk is illustrated for a leading-edge threshold discriminator; all timing circuits, whatever technique is employed, are affected in this way. Figure 8.25 illustrates the contribution to jitter from this charge requirement: amplitude variations (curves (a) and (b)) affect the time taken to deliver the fixed charge, and similarly for the rise-time variations (curves (b) and (c)). We note that $\Delta t_1 < \Delta t_2 < \Delta t_3$ and that triggering always takes place at a fixed time after the threshold is exceeded. A fast rise time in the leading edge not only helps in reducing amplitude walk but also mitigates the effects of the finite charge Δq required to trigger.

8.4.5 Contribution to jitter from noise

The two dashed curves in Fig. 8.24 represent the statistical fluctuations in the amplitude of the output signal; they can equally be interpreted as fluctuations in baseline noise, ε_n , in the PMT output, together with a contribution from the electronics. Taking these effects together, the contribution ε_t in rms jitter follows from (8.22):

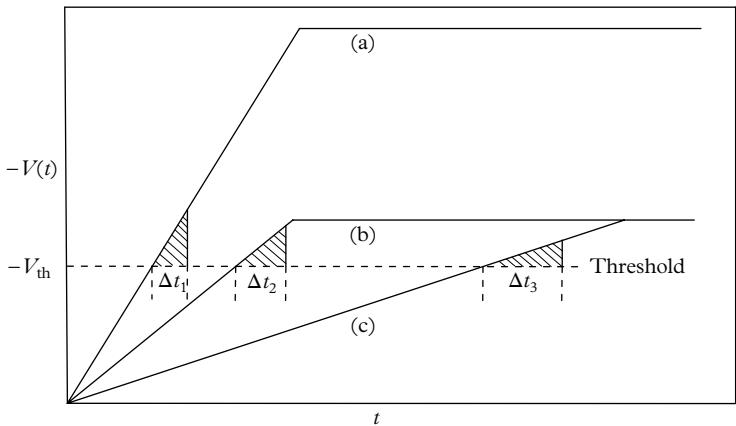


Fig. 8.25. Illustrating the contribution to jitter by an additional charge, Δq , required for triggering. Voltage pulses (a) and (b) have the same rise time but different amplitudes, while (c) has a slower leading edge. The shaded triangles represent the charge required to generate a trigger: they have the same area, but different base widths, Δt

$$\varepsilon_t = \frac{\varepsilon_n}{\frac{dV}{dt}(t)} \quad (8.25)$$

8

at $V(t) = V_{\text{th}}$. These formulae, which have quite wide application, stem from the work of Euling (1963), who in his paper warns that the exact conversion from $\Delta V(t)$ to $\Delta t(V)$ is not mathematically as straightforward as portrayed in his treatment, which is essentially heuristic. The formulae given in (8.24) to (8.25), and in the review papers of Gatti and Svelto (1966) and Poultney (1972), essentially serve as indicators of the contribution to jitter from pulse height fluctuations in q , and a contribution from noise. Equation (8.24), for example, assumes that the SER follows a normal distribution, which is doubtful in conventional PMTs. An important point that emerges in the contributions to jitter from amplitude fluctuations, charge sensitivity, and noise, is their dependence on the slope of the signal as it crosses the threshold—the higher the slope, the smaller is the contribution from these sources.

8.5 Zero crossing and CF methods

The characteristic feature of a peaked pulse is that its differential, $di(t)/dt$, has a minimum at the time of the occurrence of the peak, regardless of pulse amplitude. This is illustrated in Fig. 8.26, in which the traces shown are derived from Fig. 8.23 by differentiation. The explanation for the asymmetry of the bipolar pulses about

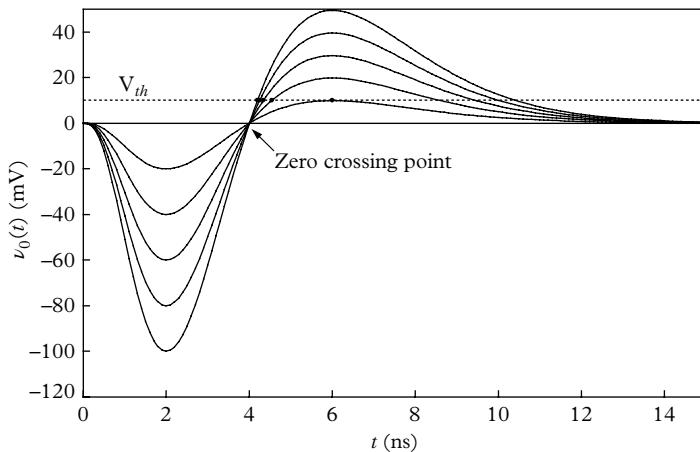


Fig. 8.26. Output voltage, $\nu_0(t)$, is derived by ‘differentiating’ the anode signal using conventional electronics circuits. All bipolar signals cross the time axis at the same position, known as the zero crossing point. Practical considerations require setting a trigger threshold at some finite value, V_{th} , as illustrated.

the zero crossing time lies in the different rise and fall times in the original pulse of Fig. 8.23. This asymmetry is particularly prominent in signals derived from inorganic scintillators because of their highly dissimilar rise and fall times, in the ratio of 10:250 in NaI(Tl), for example.

In so-called timing-filter amplifiers, the asymmetry is assuaged by a succession of differentiating and integrating filters, discussed in 14.9 and 14.10. For fast scintillator-PMT combinations, and for single-photon signals, the disparity is much reduced because the ratio of the rise time to the fall time, $t_r:t_f$, is only of the order of 1:2. Shaping amplifiers, as previously mentioned, may be used but Fig. 8.27 illustrates a rather easy way of producing a bipolar pulse from a fast unipolar one by stub clipping. The stub is a length of coaxial cable connected to the anode at one end with the other end short circuited. For example, a stub of length 12 cm of RG174 (with a delay of 5 ns/m), as shown in Fig. 8.27, produces an inverted signal at the anode at 1.2 ns after the initial signal. The output from this arrangement is the sum of the two signals with the required bipolar form. We note that the position of zero crossing, on the time axis, is common to pulses of all amplitudes—exactly as before. This circuit has many virtues, principally: it is passive, inexpensive, and easily adjusted for optimum performance. Zero crossing is sensed by a threshold discriminator set as low as the noise will allow—usually less than a few millivolts.

Achieving precise timing depends on sensing the time of zero crossing. Circuits first appeared in the 1960s, and the paper by Metz (1967) on the early use of delay line differentiation and tunnel diodes is both readable and informative. The essential feature of a tunnel diode is ultra-fast switching between two stable states,

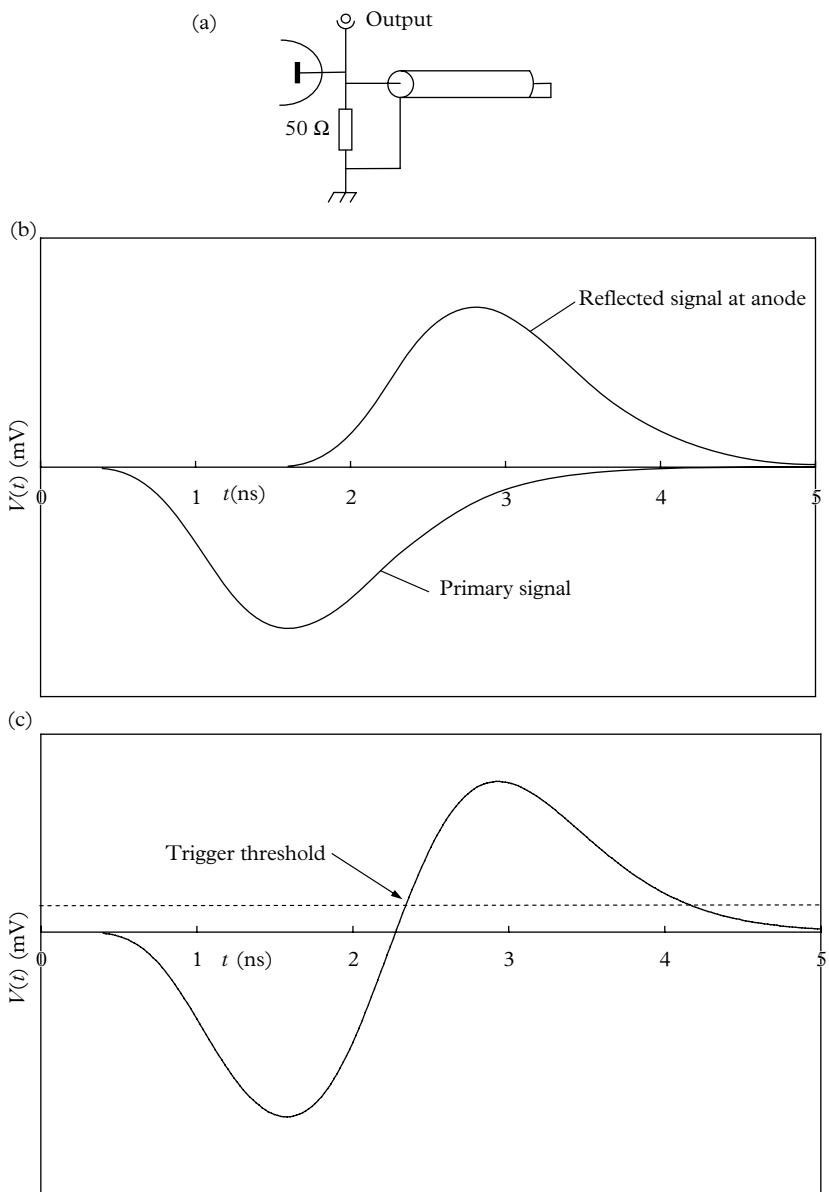


Fig. 8.27. (a, b) A simple but effective passive circuit for the generation of a bipolar pulse at the anode of a PMT. The rise time and the delay time are each 1.2 ns. (c) The signal at the summing junction.

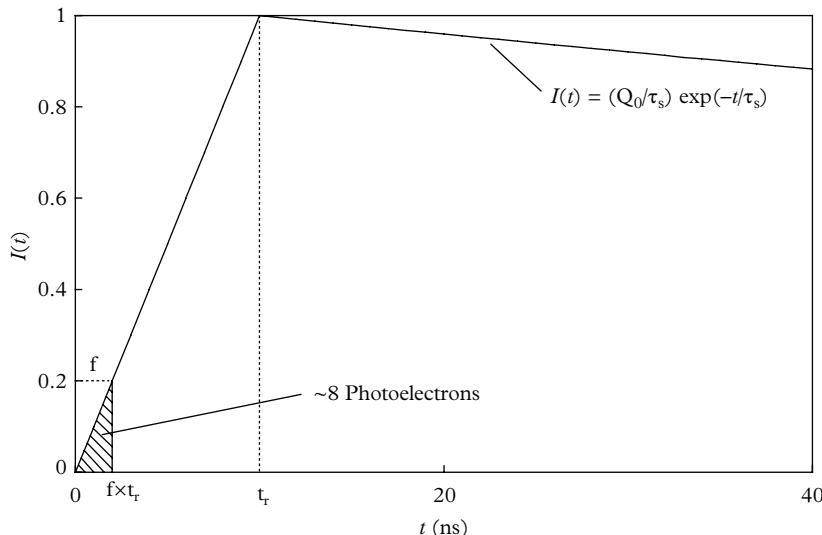


Fig. 8.28. The shaded area represents $Q(t_r) \approx 8$ pe for a NaI(Tl) pulse containing a total of 10^4 pe/MeV. The discriminator level is set at $f = 0.2$ in this example; $Q(t_r)$ depends on the ratio of the rise and fall time.

linked by a negative resistance characteristic. The tunnel diode triggers on the leading edge of a bipolar pulse, but precise timing occurs on retriggering on zero crossing. Using a fast NATON 136 scintillator and 56AVP PMTs, Metz achieved timing with ϵ of 230 ps for ^{60}Co .

8

8.5.1 Charge threshold for triggering

Timing is based on sensing the time at which the output signal exceeds a fixed threshold and, in order to understand how the various timing discriminators function, it is necessary to investigate the charge delivered up to the point of threshold, as a function of this threshold. Establishing the relationship is readily accomplished for any given signal profile, and to illustrate this we assume a NaI(Tl) pulse with a leading edge of 10 ns. All PMTs have rise times, ranging from 0.5 to 10 ns, and similarly for scintillators and Cerenkov converters. The idealized pulse in Fig. 8.28 illustrates the methodology used to determine the number of photoelectrons required to trigger a discriminator, $Q(t_r)$. If $Q(t)$ is the charge delivered up to a time t where the total charge is Q_0 , then

$$Q(t) = \int_0^t I(t) dt = Q_0/\tau_s \int_0^t \exp(-t/\tau_s) dt.$$

Specifically, if t_r is the rise time of the pulse,

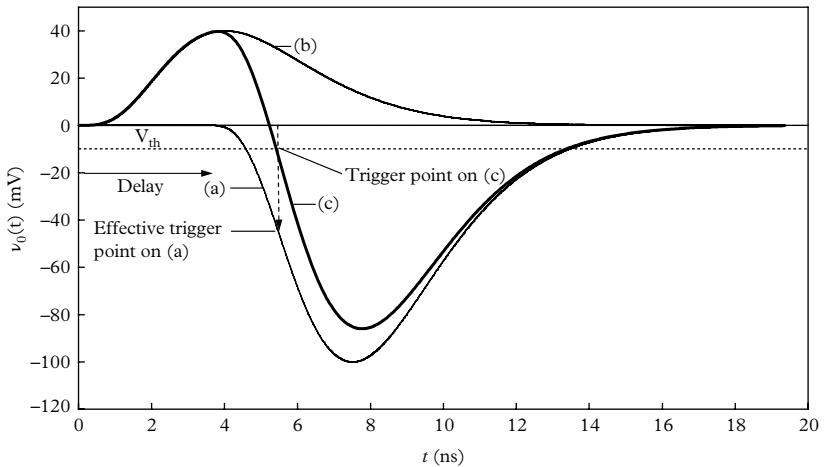


Fig. 8.29. Illustrating the CF timing technique of Gedcke and McDonald (1968). A positive-going pulse (b) is derived from a replica of the anode signal which is inverted, and attenuated by a fraction, f. This pulse, (b), is mixed with a delayed version of the anode pulse, (a), resulting in the profile (c).

$$Q(t_r) = Q_0(1 - \exp(-t_r/\tau_s)).$$

If $t_r \ll \tau_s$, which it is for NaI(Tl), then

$$Q(t_r) = \frac{1}{2} \frac{Q_0 t_r}{\tau_s}.$$

The constant of $\frac{1}{2}$ allows for the finite rise time of the signal. If a threshold is set at some fraction, f, of the maximum pulse height, then the charge delivered up to this trigger threshold is

$$Q(t_r) = \frac{1}{2} \frac{Q_0 f^2 t_r}{\tau_s}. \quad (8.26)$$

Taking the number of photoelectrons produced by NaI(Tl) as 10^4 pe/MeV, together with $f = 20\%$, and $\tau_s = 250$ ns, then $Q(t_r)$ equates to only 8 pe, or $\sim 0.1\%$ of Q_0 . This is in good agreement with the figure of 0.2 % given by Nutt *et al.* (1970), who report a rise time for their NaI(Tl) sample of 8 ns. Figure 8.28 represents a typical NaI(Tl) signal profile with a 10 ns rise time and a 250 ns decay time. The ratio of the shaded area to the total is given by (8.26); $Q(t_r)$ is small in magnitude in this case, as a result of the disparity in the rise and fall times of the pulse, for which $t_f/t_r \approx 2.5$. For fast scintillators, such as NE111, t_f/t_r is much closer to unity, and the assumption made in the derivation of (8.26) is not met; $Q(t_r)$ is now a much higher proportion of Q_0 and, for the fastest scintillators, this could amount to 10 %. In the case of fast pulses, it is expedient to determine $Q(t_r)$ by graphical means.

8.5.2 CF technique

We note with a bipolar pulse that the time at which zero crossing occurs corresponds to the peak of the undifferentiated anode signal—that is, zero slope—and that better timing would result if it could be arranged that crossing corresponded to a point of higher slope located on the leading edge of the undifferentiated signal. This is the implication of the arguments summarized in Fig. 8.29. There is another compelling reason for discriminating near the start of the leading edge, rather than on the peak of the pulse; the explanation is the statistical one presented in (8.26).

The CF timing technique, illustrated in Fig. 8.29, achieves effective leading-edge discrimination by operating on a composite signal consisting of delayed, inverted, and attenuated components. The delay is chosen to be of the order of the signal pulse rise time, for the reason given above; and f , defined in Fig. 8.28, is set within the range of 0.1 to 0.5. In the hypothetical case represented by Fig. 8.29, optimal timing is obtained with a delay of 3.5 ns and with $f=0.4$. It is reported in the literature, based on extensive experimentation with single photons and fast plastic scintillators, that the settings for these parameters are not too critical, by realizing optimal timing with $f \approx 0.2$. A family of waveforms based on Fig. 8.29 is given in Fig. 8.30.

Circuits operating on a differentiated version of the input signal, as in Fig. 8.26, are referred to as zero crossing types. Those based on mixing an attenuated and inverted signal component with a delayed version of the original pulse are of the CF type. Timing performance as a function of pulse height is read off the family of curves relevant to each technique (Fig. 8.31). In the case of the zero crossing type

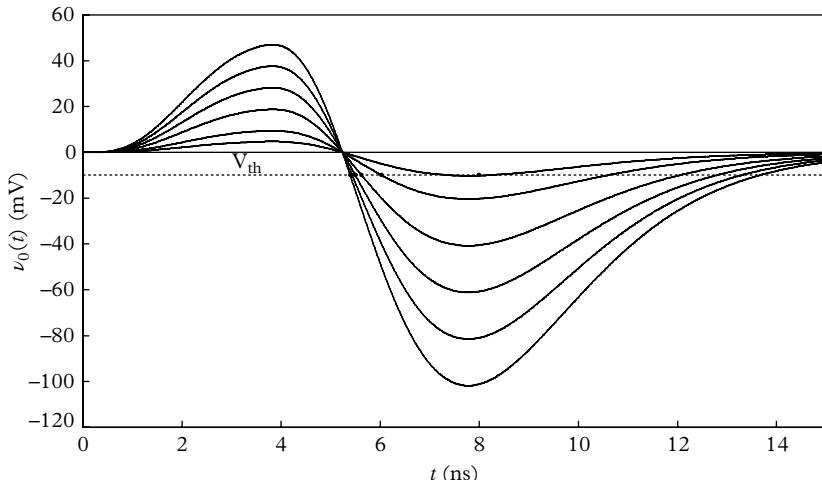


Fig. 8.30. CF waveforms derived from the unipolar set in Fig. 8.23, using the method of Fig. 8.29, with $f = 0.4$, and a delay of 3.5 ns.

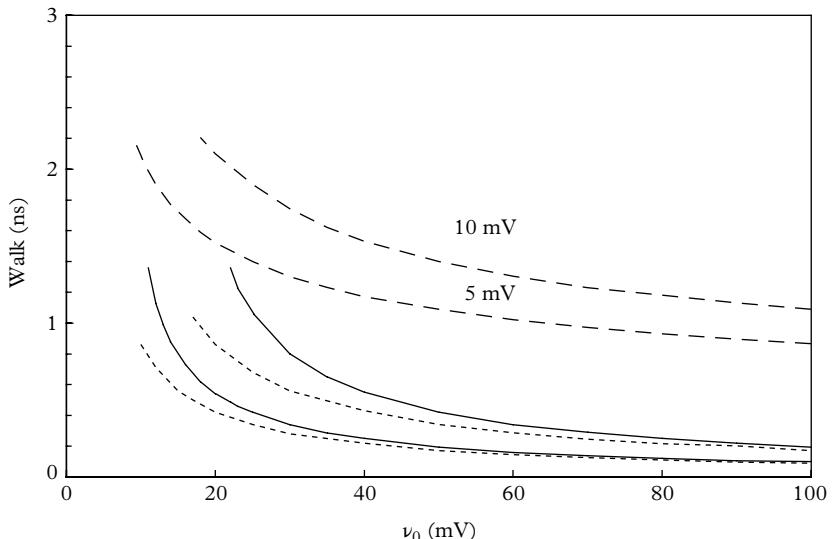


Fig. 8.31. Walk as a function of peak pulse height for two discriminator settings, $V_{\text{th}} = 5 \text{ mV}$ and 10 mV . The upper pair of curves refers to threshold detection, the next set (solid lines) is for the zero crossing technique, and the lowest dashed pair describes CF discrimination. All curves decrease monotonically to the right, ultimately approaching zero walk.

8

and the CF type, the time origin is taken at the crossing point, as the delay is of no consequence. The curves in Fig. 8.31 suggest that a dynamic range of 100:1 is feasible with these types of discriminator, as will now be justified. We observe that the walk rises very rapidly for a pulse just above threshold, but it flattens once $\nu_0(t)$ exceeds about 50 mV in amplitude. Taking this signal level as the starting point of the range, and noting that a fast PMT can deliver a pulse of up to ~ 5 volts, equivalent to 100 mA into 50Ω , such a PMT would provide the required 100:1 dynamic range. At a threshold of 5 mV, the CF walk ranges from about 200 ps to near zero over this dynamic range. In practical circuits, the timing trigger threshold is close to zero but, as discussed in 8.4.4, a finite charge is necessary to fire the discriminator; this charge can be equated to an effective threshold of V_{th} . This threshold is close to the noise but in commercial units a coincidence is demanded with a variable amplitude threshold discriminator (energy selector) which prevents spurious triggering—this is explained in Fig. 8.32.

As noted previously with reference to Fig. 8.26, the zero crossing point for differentiated signals corresponds to the peak of the anode signal for all pulse shapes. Referring to fast pulses with $t_r \approx t_f$, zero crossing occurs after a charge within the range of $Q_0/2$ to $Q_0/10$ has been delivered by the PMT—it is exactly $Q_0/2$ only for a perfectly symmetrical anode pulse, and this is certainly not the case for inorganic scintillators. For the CF discriminator (CFD), the zero crossing

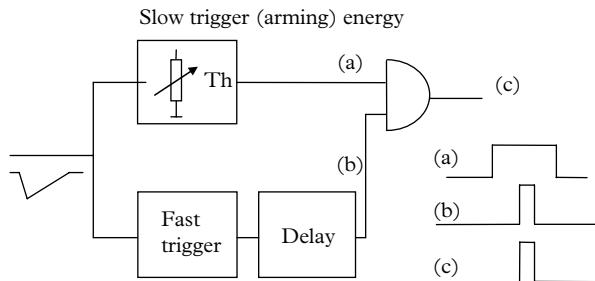


Fig. 8.32. A threshold discriminator typically consists of two parallel channels. The signal from the slow channel, indicated by (a), always arrives at the AND gate ahead of the fast trigger signal, albeit with considerable time walk. The fast trigger may include a CF discriminator, a zero crossing discriminator, or a threshold discriminator.

point relates a point on the leading edge of the anode pulse, as illustrated in Fig. 8.29, and consequently to a much lower $Q(t_r)$ value than $Q_0/2$. These considerations, naturally, also apply to a leading-edge threshold discriminator.

8.6 Measurement schematics

8

A common prerequisite of the different discrimination techniques is the need to set a low discrimination level. The requirement is to accept only those PMT pulses that fall within a prescribed size band: a facility all good commercial instruments provide. Selection must be made in an independent and parallel channel, referred to as a ‘slow’ or ‘energy’ discriminator, as illustrated in Fig. 8.32. The signal feeding the slow channel is usually smoothed and shaped, and detection occurs whenever the set threshold is exceeded. The fast trigger circuit may be any of the three discriminator types already discussed. The fast threshold level adjust is not normally accessible to the user but is factory set at a very low level close to or even within the noise level of the electronics. The schematic for a CF type is given in Fig. 8.33, with the delay provided by a short length of coaxial cable with a delay of the order of 5 ns/m.

The units described in Figs 8.32 and 8.33 give the essential operating principles of discriminators used in timing. Figures 8.34 to 8.36 show examples of timing apparatus, but more detail is given by Nicholson (1974), and also in an ORTEC application note by Gedcke and Williams (1968). This pair of references is particularly recommended for more detailed information covering all aspects of timing.

Single-photon timing distributions serve as a measure of the performance and capability of a PMT. Multi-photon timing distributions are particularly relevant to scintillation spectroscopy, for example, where the combined timing performance

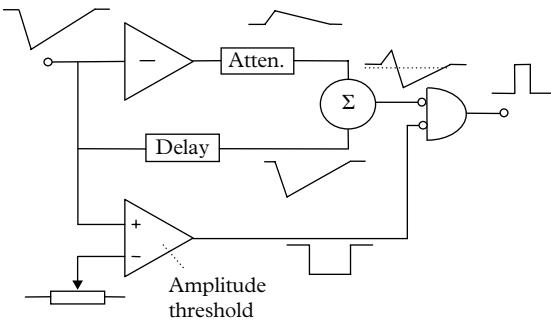


Fig. 8.33. The essentials of a CF discriminator embodying two parallel channels, as in the previous example.

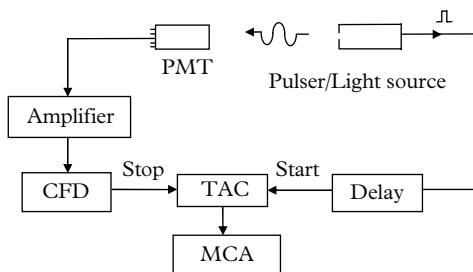


Fig. 8.34. Conceptual electronics configuration for single-photoelectron time resolution measurements using an electrical start pulse. Experimental details such as the sample chamber, filters, and the monochromator have been omitted.

of the detector and the electronics over a specified energy range is of primary concern. Configurations used for experimental investigations, involving timing of single and multi-photoelectron events, are essentially the same as those that are suitable for investigating the performance of the PMT itself (Poultney 1972). In addition, the stability of the light source, electronics, and calibration referred to as system diagnostics by Poultney may also be investigated. The illustrative examples given in Figs 8.34 to 8.36 describe the principles of the configurations in common use.

The schematics in Figs 8.34 and 8.35 were used by Leskovar and Lo (1975) and by Moszyński and Vacher (1977) over many years. These two scientific groups established themselves as centres of excellence for the evaluation of new PMTs and scintillators. The essential timing unit is the time-to-amplitude converter (TAC), which measures the time interval between two logic pulses—the ‘start’ pulse and the ‘stop’ pulses. The light source is often an LED or laser diode, which also provides an electrical ‘start’ pulse. The pulse generated by the PMT, if of sufficient amplitude to trigger the discriminator, provides the stop pulse. Accumulating many ‘stop–start’ events generates a timing spectrum for

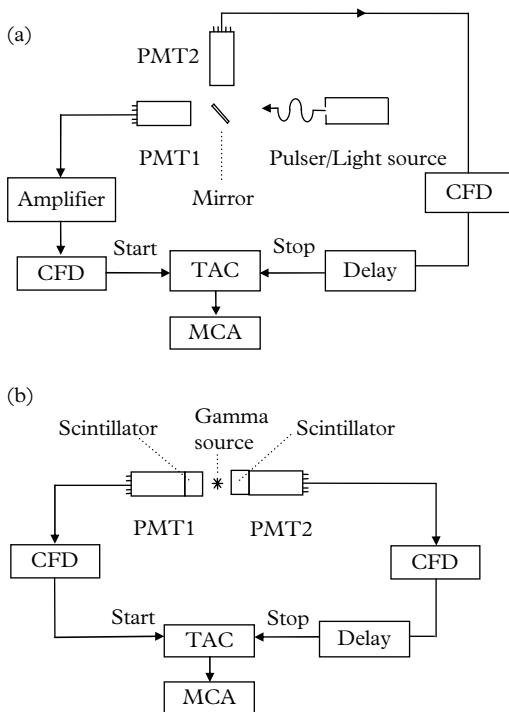


Fig. 8.35. (a) A configuration similar to the one shown in the previous figure, except the start pulse is derived from a second PMT. Note the use of a plane glass mirror to direct predominantly single photons to the stop channel. (b) The pulser is replaced by scintillators and, if the scintillators and PMTs are of the same type, the timing dispersion of each detector can be taken as $1/\sqrt{2}$ times that measured for the combination.

storage in an MCA, for example. The timescale may be adjusted, within the range of 1 to 100 ps/channel, to suit the width of the timing spectrum. A delay unit compensates for the combined propagation time through the PMT and the CFD, which can extend out to 50 ns. Fast, commercial light sources such as the Hamamatsu Picosecond Light Pulser PLP-10 and the IBH nanoLED generate up to 10^6 photons per flash of duration 70 ps, in the case of the PLP-10. Low-repetition-rate, picosecond, mode-locked lasers, such as those used for LIDAR, deliver up to 10^{13} photons per flash but at a low repetition rate. Light sources made in the laboratory, based on avalanche transistors and LEDs, provide pulses of sufficient speed and light output for PMT calibration purposes. It is necessary to attenuate the light where single-photon excitation applies. This arrangement is inherently inefficient since it is a necessary condition that most start pulses should fail to produce an output. A success rate of between 1 and 10 % is usually employed to ensure essentially single-photon excitation with negligible plurality. Some experimenters exchange the stop and start channels, compensating

with additional delay, in which case the direction of time in the MCA display is reversed. The reason for this is to reduce dead time caused by the inefficient single-photon generation process. With the start and stop functions reversed, only the successes generate a start pulse.

The timing arrangement shown in Fig. 8.35(a) uses a fast PMT for generating a start pulse of negligible jitter by taking advantage of the improvement in timing fidelity with the number of photoelectrons in the signal; PMT2 in the stop channel operates in the single-photon mode. In some experimental configurations, it is found that better performance results with a leading-edge discriminator in the ‘start’ channel, rather than a CFD. The arrangement in Fig. 8.35(b) is sometimes operated with identical channels but other variants are possible with combinations of organic and inorganic scintillators (see Gedcke and Williams 1968).

The configuration of Fig. 8.35(a) is easily adapted for the measurement of the intrinsic time dispersion characteristic of a particular PMT. This is done by replacing the pulser/light source with a Cerenkov converter including a low activity ^{90}Sr source of relativistic electrons. A small-volume light source is closely coupled to PMT1, ensuring that most of the light generated per event is collected by this PMT. There is an occasional coincidence of a single-photon event in PMT2 with PMT1, and the spectrum of these events provides a single-photon timing distribution. An effective arrangement consists of a small cylinder of transparent material, with one face in contact with the start PMT; the stop PMT is mounted at right angles to the start PMT at a distance of a few centimetres, and a notch cut into the cylinder delivers single photons to this PMT. The acquisition time is necessarily long but cost is minimal, and long-term stability is assured. Rearranging the disposition of the light source so that both PMTs are approximately equally illuminated with single photons produces a joint distribution for the pair of PMTs. If they are of the same type, or are known to have similar timing characteristics, then a reasonable measure of the individual performance is afforded by scaling the fwhm

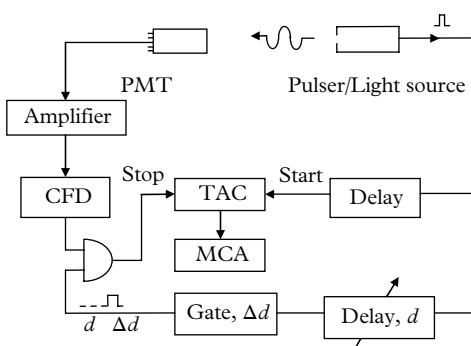


Fig. 8.36. A timing system specifically for determining late pulse and afterpulse rates. The recorded events are those for which the stop pulse occurs within a time of Δd after the start. Measurement periods in this delayed coincidence arrangement tend to be long.

by $1/\sqrt{2}$. An ideal timing distribution is one that is not statistically limited, thereby revealing small artefacts and the presence of any fine structure. Note that a total of about 100 events is sufficient to arrive at a reasonable figure for jitter.

8.7 Multiple statistical sources

The essential electron-optics problem, evident in Fig. 8.11, is one of efficiently coupling the axial symmetry of the cathode and its associated focussing elements to the planar geometry of the dynodes. Secondaries from the first dynode must exit at a right angle with respect to the direction of the primary electrons, in order to reach d_2 . An undesirable complication in some designs is d_3 when chosen smaller in area than the common dynodes, which start at d_4 . The effect on timing performance can be mitigated by increasing V_{k-d1} , thereby reducing the contribution derived from the finite photoemission energy, tighter focussing, and reduced spot size. Contributions from the remaining pairs of dynodes are smaller because of compensation discussed in 8.2.2.

Minimizing the variation in the time interval between photoelectron emission and the generation of a corresponding anode signal is fundamental to achieving precise timing. This statistical property is illustrated in the form of a histogram (see Fig. 8.22) of arrival times, and the preferred figure of merit is the fwhm height, t_{fwhm} . The jitter specification of a device, usually quoted in nanoseconds or sub-nanoseconds, quantifies the attainable precision in timing for single-photoelectron excitation. Jitter also affects the pulse shape of single- and multi-photoelectron-initiated signals, as seen at the anode. Pulse shapes are described by three parameters: rise time, t_r ; *fwhm*; and fall time, t_f . Rise time is the most relevant of the three parameters to timing applications. Examination of manufacturers' and published data shows a relationship between t_{fwhm} and t_r , although the correlation is rather weak. Clearly, the time response of a PMT is particularly relevant when observing fast transient phenomena, but it is also important in minimizing overlap when high count rates are encountered.

8.7.1 Single-photoelectron output signatures

Following Gatti and Svelto (1966), a PMT output signal may be represented by a function, SETR, embodying the area of the pulse (essentially the gain, $\langle g \rangle$, since we are concerned with single photoelectrons) and its width; T is defined as the interval between emission of a photoelectron and the centroid of the corresponding output pulse at the anode. The width of the pulse is defined by λ^2 , the inverse of the variance about the centroid (see Fig. 8.20). The mean value of $\langle g \rangle$, measured in electrons, for $n - 1$ dynodes, is $(\delta_1, \delta_2, \dots, \delta_{n-1})$ and, if the stage gains are all equal, $g = \delta^{n-1}$. The mean value of T is

$$T = t_0 + t_1 + t_2 + \dots + t_{n-1}, \quad (8.27)$$

where t_0 is the mean transit time from the cathode to the first dynode, and t_n is the time taken between the last dynode and the anode—the latter is often ignored as it is small. The mean value of λ^2 is obtained by adding variances in quadrature, thus obtaining

$$\lambda^2 = \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_{n-1}^2, \quad (8.28)$$

where ε_n^2 is the variance of the pdf for flights between dynodes n and $n+1$. There is a subtle point to note regarding the omission of ε_0^2 in (8.28): remember that we are dealing with single-photoelectron emission, so the signal-forming process starts at the first dynode; consequently, while t_0 contributes to T and the variance in T , its related variable, ε_0^2 , does not contribute to the variance in the SETR. If all ε_n 's are equal and denoted by ε_{d-d} ,

$$\lambda^2 = (n-1)\varepsilon_{d-d}^2. \quad (8.29)$$

8.7.2 Photoelectron timing variance

Gatti and Svelto (1966) published a review of resolving time in scintillation detectors as a critique of the then current theories, including methods for measuring dispersion. They considered the following list of sources responsible for transit-time dispersion:

8

- (1) flight of a photoelectron from the photocathode to the first dynode
- (2) the path between the first and the second dynode
- (3) transitions between pairs of the remaining dynodes
- (4) dispersion introduced by the collection of charge by the anode

The omission of a contribution from photoemission lag is notable by its absence, but is easily explained as the pulsed light sources and PMTs available at that time were too slow to reveal this effect. Furthermore, scintillators emit mostly in the blue or blue-green region of the spectrum, where time lag is minimal. Nowadays, photoemission lag is a consideration, especially at long wavelengths, with the (NEA) III–V photocathodes. PMT users may consider that (1) to (4) are the concern of the designer and not the user. However, knowledge of these phenomena helps in selecting the most appropriate detector, allowing for a compromise between financial constraints and sufficient quality of operation. As may be anticipated, there is a relationship between timing performance and cost. In any case, having made the selection there remains the challenge of optimizing performance. Gatti and Svelto (1966) provide the statistical framework for the overall transit-time dispersion, the variance, ε_{PMT}^2 . The first contribution, moving in the direction of cathode to anode, is that of dispersion incurred in the transit between

the photocathode and first dynode, ϵ_{k-d1}^2 . Assuming full and even photocathode illumination takes account of the different trajectories followed by photoelectrons, as discussed in 8.1. Path length and hence time taken to reach the first dynode varies, depending on the point of initiation on the photocathode, the initial energy, V_0 , of a photoelectron, its launch angle, and its landing position on d_1 . Fast PMTs invariably include a plano-concave window, in preference to a plano-planar one, to reduce the spread in path lengths. The second contribution, ϵ_{d-d}^2 , stems from the fluctuations introduced by the multiplier, but given the major role of the first dynode, it is customary to treat this dispersion, ϵ_{d1-d2}^2 , separately from the remainder—consequently, the pairs of common dynodes are each assigned the dispersion ϵ_{d-d}^2 . The overall time dispersion, ϵ_{PMT}^2 , is a combination of contributions from all these sources, generated by summing the variances; thus,

$$\epsilon_{PMT}^2 = \epsilon_{k-d1}^2 + \epsilon_{d1-d2}^2 + \epsilon_{d-d}^2 + \epsilon_{d-a}^2.$$

The first component, ϵ_{k-d1}^2 , accounts for contributions that involve the photocathode and first dynode, as described in 8.1. Positional dependence, according to de la Barre (1973), causes deterioration in time resolution by an additional 400 ps in moving the light source from the axis to the periphery of an XP2020 window. The latter is a statistical effect and related to wavelength. Sipp *et al.* (1976) show that time spread reduces by ~30 % at 500 nm compared with that for 350 nm, while Moszyński and Vacher (1977) report a similar improvement between 570 nm and 790 nm.

The second component, ϵ_{d1-d2}^2 , refers to the variation in electron transit time between the first and second dynode. This is treated differently from the common interdynode contributions, ϵ_{d-d}^2 , since it is customary to operate with V_{k-d1} set to twice or three times an interdynode voltage and hence with higher gain, δ_1 .

The third component, ϵ_{d-d}^2 , is the transit-time variance between adjacent common dynodes—all assumed to have a common stage gain of δ .

The fourth component, ϵ_{d-a}^2 , refers to the anode and last dynode. A typical reflex anode is spaced about 1 mm from the last dynode, and the contribution ϵ_{d-a}^2 is small.

Multiplier noise comprises the combined contributions $\epsilon_{d1-d2}^2 + \epsilon_{d-d}^2$, suitably weighted, in the manner of (4.46); thus,

$$\begin{aligned} \epsilon_D^2 &= \frac{\epsilon_{d1-d2}^2}{\delta_1} + \epsilon_{d-d}^2 \left(\frac{1}{\delta_1 \delta} + \frac{1}{\delta_1 \delta^2} + \frac{1}{\delta_1 \delta^3} + \dots + \frac{1}{\delta_1 \delta^{k-1}} \right) \\ &= \frac{\epsilon_{d1-d2}^2}{\delta_1} + \frac{\epsilon_{d-d}^2}{\delta_1 \delta} \left(1 + \frac{1}{\delta} + \frac{1}{\delta^2} + \dots + \frac{1}{\delta_1 \delta^{k-2}} \right) \\ &= \frac{\epsilon_{d1-d2}^2}{\delta_1} + \frac{\epsilon_{d-d}^2}{\delta_1 \delta} \frac{\delta}{(\delta - 1)} \\ &= \frac{\epsilon_{d1-d2}^2}{\delta_1} + \frac{\epsilon_{d-d}^2}{\delta_1 (\delta - 1)}. \end{aligned} \quad (8.30)$$

Combining all sources of variance leads to the time dispersion at the anode:

$$\epsilon_{\text{PMT}}^2 = \epsilon_{k-d}^2 + \frac{\epsilon_{d1-d2}^2}{\delta_1} + \frac{\epsilon_{d-d}^2}{\delta_1(\delta_{d1-d2} - 1)} + \epsilon_{d-a}^2. \quad (8.31)$$

Equation (8.31) originated from Piétri (1973), while Leskovar and Lo (1975) and Moszyński and Vacher (1977) include the SER noise factor for gain, $(1 + \text{var}/\delta^2) = (1 + 1/\delta)$, in the ϵ_{d1-d2}^2 and ϵ_{d-d}^2 contributions. The equation gives the total time dispersion and the relative contributions from individual sources, but requires numerical values for the four variances. Electron-optical modelling is required for this purpose or, to achieve an order of magnitude indication of the contributions from the four terms, we make use of (8.9), proceeding as follows. The transit time, T , may be roughly expressed in terms of individual transit times for a 12-stage PMT, such as an XP2020; thus,

$$\begin{aligned} T &= t_{k-d1} + 12t_{d-d} + t_{d-a} \\ &= 3t_{d-d} + 12t_{d-d} + t_{d-d}. \end{aligned}$$

According to the manufacturers, T is 30 ns, from which we deduce the interdynode transit time t_{d-d} as $\sim T/16 = 1.9$ ns. The standard deviations ϵ_{k-d1} , ϵ_{d1-d2} , ϵ_{d-d} , and ϵ_{d-a} are not readily available but an order of magnitude deduction is possible from (8.9) by assuming that Δt corresponds to one standard deviation; V_0 is assumed to be 1 eV for photoelectrons, and 5 eV for secondary electrons: $\delta_{d1} = 6$; $\delta_{d-d} = 4$; $t_{k-d1} = 5$ ns; and $V_{k-d1} = 300$ V. In addition, $V(s_0)$ is taken as 200 V for the interdynode voltages because of the accelerating effect of d_{n+2} on d_n (see Fig. 8.11). The transit time between d_1 and d_2 is taken as twice that for common dynodes to allow for the higher spacing and poor geometric compensation. The transit time between the last dynode and the anode is minimal because reflex dynodes are characterized by a spacing of only ~ 1 mm. Consequently,

$$\begin{aligned} \Delta t_{k-d1} &= \left(V_0/V(s_0) \right)^{1/2} t_0 = (1/300)^{1/2} \times 5 \text{ ns} = 0.29 \text{ ns}, \\ \Delta t_{n,n+1} &= (1/200)^{1/2} \times 1.9 \text{ ns} = 0.30 \text{ ns}, \\ \Delta t_{d1-d2} &= 2 \times \Delta t_{n,n+1} = 0.6 \text{ ns}, \\ \Delta t_{d12-a} &= (5/100)^{1/2} \times 0.5 \text{ ns} = 0.11 \text{ ns}, \\ \epsilon_{\text{PMT}}^2 &= (0.29)^2 + \frac{(0.6)^2}{6} + \frac{(0.3)^2}{6 \times 3} + 0.11^2 = 0.16 \text{ ns}, \\ \epsilon_{\text{PMT}} &= 0.40 \text{ ns}. \end{aligned}$$

The overall standard deviation is 0.40 ns, or $2.35 \times 0.40 = 0.94$ ns, fwhm.

The actual value quoted by the manufacturer for an XP2020 is $\sigma = 0.25$ ns. The contribution from the term ϵ_{d12-a}^2 in (8.30) has been nicely demonstrated by Moszyński (1993b), who reports on a version of the XP2020—the XP2020UR-M. This variant, discussed in 8.2.3, includes a grid mounted between the anode mesh and the penultimate dynode but disposed close to the anode to intercept the induced current produced by electrons approaching the anode. This effect is of opposite polarity to the main signal and smaller in amplitude (see e.g. de Vries and

Kelling 1988). Moszyński (1993b) quotes a timing improvement of 14 % for a range of multi-photoelectron pulses from fast scintillators, and the same improvement from a slow NaI(Tl) scintillator.

Assumptions made in arriving at (8.31) and the uncertainty in ϵ parameters implies that the formula can only provide an order of magnitude prediction for time dispersion. Its main virtue is to reveal the dependency on V_{k-d_1} , $V_{d_1-d_2}$, and δ_1 . Equation (8.31) predicts the total contribution to dispersion from individual PMT ϵ parameters, and it is immediately obvious that a good SER and hence high d_1 gain are important in lowering dispersion. The equation can only be used to predict timing performance if the numerical values for ϵ terms are known. Timing performance of any existing or proposed PMT can be obtained by computer simulation: the more advanced, proprietary 3-D electron-optics programs, which include a Monte Carlo simulation of launch energy and emission angle, are able to provide the necessary data from which numerical values for the ϵ terms can be calculated. One may be tempted to assume that a PMT with small λ will also have low jitter, but this is not always so.

The immediate insight following from (8.29) and (8.31) is that, the fewer the number of stages of gain, the lower is the pulse width dispersion. RCA exploited this relationship in their offering of the fast performance, five-stage C31024 PMT, with GaP(Cs) dynodes throughout. Equation (8.29) indicates that λ^2 for the 5-stage C31024 will be half that for the 12-stage RCA8850, from which it derives. In practice, there is further improvement in timing because the interdynode potentials in the 5-stage PMT are twice those for the 12-stage 8850, leading to a further reduction in ϵ_{d-d} . This was confirmed by Leskovar and Lo (1978), who reported on the performance of the 8850 and the C31024 as follows: $t_r=2.4$; $\lambda=2.1$; $\epsilon_{PMT}=0.45 \times fwhm$; and $t_r=0.8$; $\lambda=0.4$; and $\epsilon_{PMT}=0.40 \times fwhm$ (all in nanoseconds). There is more information in (8.29) apart from being able to predict the SETR: it actually describes the stage-by-stage evolution in shape of the signal as it progresses towards the anode. The signal gets broader, consistent with the portrayal shown in Fig. 8.20.

8.8 Photon arrival statistics

Statistically, there are many light sources, $F(t)$, to consider but it turns out that most PMT applications fall into two categories: in the first, the light-emission profile is exponential, involving multiple photons per event (scintillation and excitation processes); in the second, light detection is single-photon based (e.g. time domain fluorescence spectroscopy). The pdf $P(t, \lambda, Q|R_m)$ for photoelectrons has the same form as that for photons, except for normalization with respect to QE. This equivalence is valid, provided there is negligible delay between the detection of a photon and the corresponding emission of a photoelectron.

Following the discussion in 8.1.1, any lag in emission, for wavelengths characteristic of scintillator light, will be of the order of picoseconds, and hence the contribution may be ignored for present purposes.

Variance in the arrival time of photons produced by a source characterized by an exponential decay was derived in 4.16.3 and is repeated in (8.32):

$$\begin{aligned} \text{var}(Q|R_m) = & \frac{Q\tau_s^2}{R_m^2} \left[1 + \frac{2(Q+1)}{R_m} + \frac{(Q+1)(6Q+11)}{R_m^2} \right. \\ & \left. + \frac{(Q+1)(Q+2)(2Q+5)}{R_m^3} \right]. \end{aligned} \quad (8.32)$$

The dispersion for exponential signals is given in Fig. 8.37 for three inputs R_m : 100, 1000, and 5000 pe. The abscissa refers to the threshold in terms of the fractional ratio Q/R_m . Contributions from the individual terms in square brackets in (8.32) arise predominantly from the first term (Fig. 8.38).

It is evident in (8.32) that the coefficient $Q\tau_s^2/R_m^2$ predicts that precise resolving time can be achieved, even with slow scintillators, provided that R_m is of sufficient magnitude. The variable terms within the square brackets make a small contribution to the variance, provided Q is less than about 4 % of R_m , as is evident in Fig. 8.37. In this instance,

$$\text{var} \approx \frac{Q}{R_m} \times \tau_s^2 / R_m, \quad (8.33)$$

8

and, since Q/R_m is constant for CF timing, the standard deviation should therefore vary as $\tau_s / R_m^{1/2}$. Equation (8.32), illustrated in Figs 4.16 and 4.17, predicts that optimal timing (lowest variance) is achieved based on detecting the arrival of the first

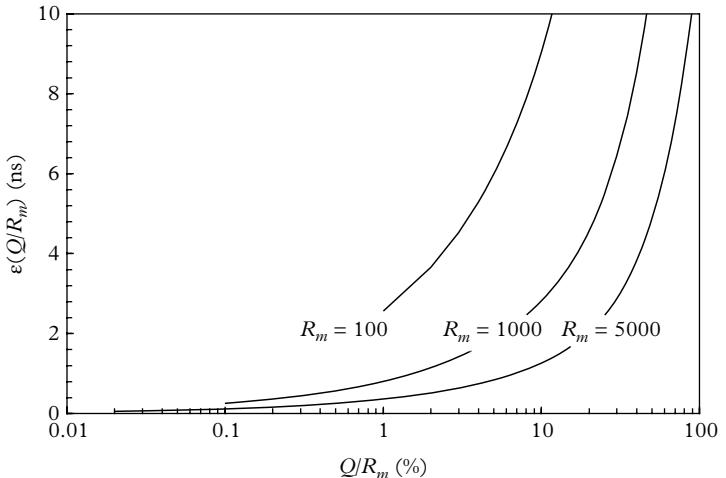


Fig. 8.37. The standard deviation for emission times derived from (8.32) for $\tau_s = 250$ ns.

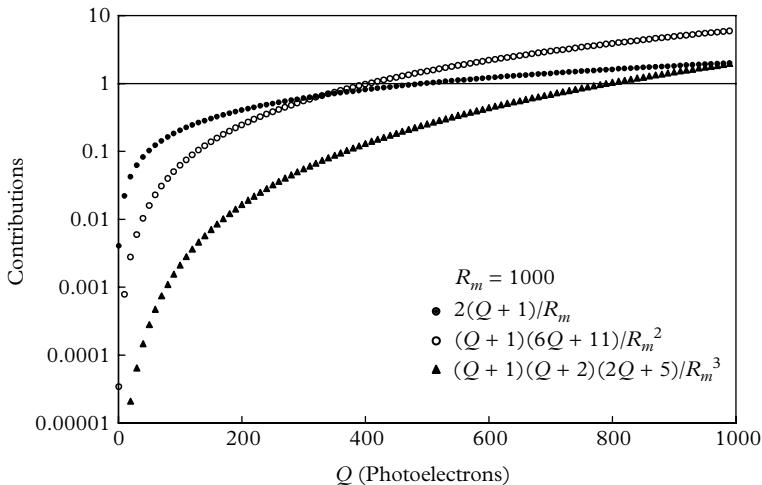


Fig. 8.38. Contributions to (8.32) from the four terms in square brackets. The contributions from the three variable terms comprise less than 10 % for Q less than 40, for example.

photon. This applies to the ideal case with R fixed and for R_m Poisson distributed. Note in Fig. 4.16 that randomizing the light source has considerable influence on variance for low R_m but, as the number of photoelectrons increases, $\text{var}(Q|R_m)$ and $\text{var}(Q|R)$ tend to converge at low Q values. It is instructive to plot the square root of (8.32) for a range of R_m values, as in Fig. 8.37, to reveal a strong dependence on Q/R_m particularly for low values of R_m . NaI(Tl) is a slow scintillator but, for gammas of sufficiently high energy (e.g. ^{60}Co), a timing of $\epsilon=160$ ps has been achieved by Moszyński (1993a). This paper, together with that of Braunsfurth and Körner (1965), illustrates the $1/\sqrt{E}$ dependence but does not support the first-photon prediction for best timing. The latter authors find that $Q \approx 75$ gives best timing for 1250 keV gammas in their particular experimental arrangement.

It should now be clear why discriminators based on the CF principle are the preferred choice for fast-timing applications involving single-photoelectron pulses, and a wide range of scintillators. On the other hand, the zero crossing technique is applied to differentiated waveforms and consequently discriminates at high Q values corresponding to the peak of the original signal, while the CF types trigger on the rising edge. The Q -value trigger point for zero crossing corresponds to $Q \approx R/2$ for symmetrical pulses, but for NaI(Tl) the peak emission rate occurs at less than 20 ns from the start (see Fig. 8.19) and the effective Q value is correspondingly low at around 0.1 %; zero crossing and CF discriminators give similar performances. The significance of a contribution from transit-time jitter, measured in sub-nanoseconds, is not obvious with slow scintillators: intuitively, one might expect an indiscernible contribution from a PMT in a photoemission process characterized by a 250 ns time constant. Gatti and Svelto (1964) explain

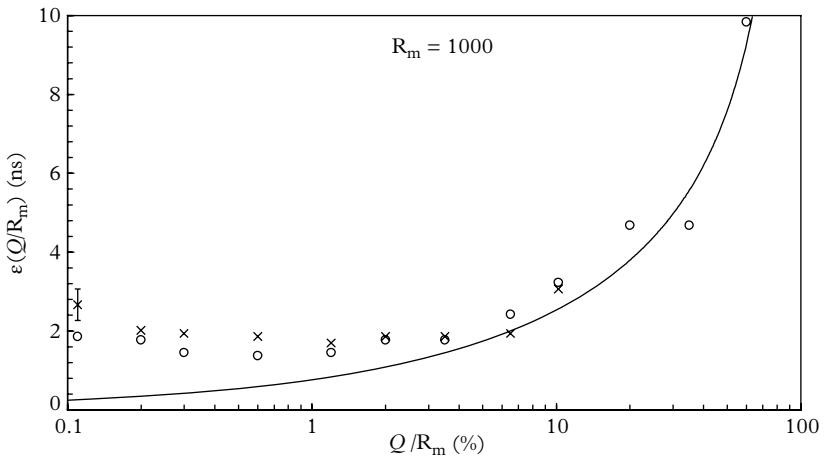


Fig. 8.39. Timing distributions for $R_m = 1000$, where $\epsilon_{PMT} = 2.5$ ns for the open circles (\circ), and 5 ns for the crosses (\times). The scintillator is NaI(Tl), with $\tau_s = 250$ ns. A single error bar gives an indication of the simulation precision; this is limited primarily by the demands of long run times. The solid line is taken from Fig. 8.37.

why this thinking is wrong. Monte Carlo simulations given in Fig. 8.39, made by the present author, indicate an optimal setting for Q/R_m at 1 % for NaI(Tl), although the precise setting is not too critical. Time taken to accumulate sufficient charge to activate the discriminator is also a determining factor, whatever type of discriminator is employed. Timing fidelity also depends on the rate of arrival of charge at the anode, which is proportional to R/t_r , a figure of merit sometimes quoted. Although t_r is about 10 ns for NaI(Tl), compared with picoseconds for undoped organic scintillators, R values are generally high.

8.9 Timing measurements

Following the invention of the CFD by Gedcke and Williams (1968), two centres of excellence developed—one in the USA and the other in Europe. Leskovar and Lo (1972) of LBL, Berkeley, initially made single-photoelectron measurements using zero crossing discrimination. Time spreads for the 8850 and the C31024, the preferred types for fast timing, are 0.64 and 0.57 ns, respectively. Six years later, using CF timing, they attained 0.48 ns and 0.42 ns, respectively, for the same PMT types (Leskovar and Lo 1978). Measurements for restricted photocathode illumination showed an improvement by a factor of 1.4 in these parameters. Time resolution over the range 1 to 10^4 pe follows a slope of $Q^{-1/2}$ and is consistent with measurements on sodium iodide, by other investigators.

M Moszyński, from the Soltan Institute for Nuclear Studies, Świerk, Poland, established a facility similar to that of LBL for investigating single-photoelectron timing, and performance with scintillating organic and inorganic crystals. Collaboration with Philips saw the development of an improved XP2020 range, with the introduction of the XP2020UR and the XP2020UR-M giving timing in the region of 200 ps, compared with 270 ps for the XP2020. Three technical publications by Moszyński (1993a, b, 2004) provide a comprehensive collection of timing measurements for plastic scintillator, organic, and inorganic crystals. Of particular interest is the fast response from LaBr_3 (Ce), which exceeds the performance of $\text{NaI}(\text{Tl})$ by virtue of higher light output and faster decay.

The measurements of Derenzo and Moses are worthy of mention, for two reasons. The equipment devised provided unequalled time resolution through the inclusion of a MCP, in preference to a PMT, as the key element. This group has characterized possibly the most extensive range of scintillator pulse rise and fall times (see 8.3.5).

Timing performance of any PMT depends on its design, the details of the divider network, and the voltage at which it is operated. Making meaningful comparisons of measurements is often difficult when considering the results of different workers, even when the same PMT type is considered. In the investigations of Leskovar and Lo (1978), PMTs were operated up to 4000 V, whereas Moszyński (1993a) chose 2000 V for the same PMT type; we have seen in (8.9) that jitter has a $1/\sqrt{V(s_0)}$ dependence, and similarly there is also a dependence on HV. The region of the photocathode illuminated also affects jitter: it is least for central exposure. Such considerations lead to uncertainties when comparing measurements by various authors.

As pointed out by Moszyński (1993a), there is a contribution to jitter of 33 ps/cm (in air) associated with light collection from a variable point of interaction in a scintillator. There is a contribution to jitter of 50 ps/cm of path length, for NE111 (refractive index 1.5), and approaching 70 ps/cm for the optically dense scintillators, with refractive index ~ 2 . Derenzo *et al.* (2000) show graphically the effect of including all internally reflected rays in a 10 mm cube of PbWO_4 , discussed in 8.3.5. This is achieved by the addition of white reflector, which, when replaced by black paint, changes the rise time from 400 ps to 100 ps. It is clear that timing measurements concerning scintillators always include a contribution due to the transit time of the photons to the cathode. Variable optical transit time has the effect of slowing the rise time of light emission which, together with intrinsic rise-time effects in scintillators, leads to degradation in timing. As we saw in 8.4.4, a finite charge is required to trigger a discriminator; the rate of accumulation of this charge is critical to timing, and signal rise time can only slow its delivery.

An accepted method for determining the jitter of a PMT–scintillator combination is to use two such detectors and a source emitting coincident gammas, such as ^{22}Na or ^{60}Co . A simplified version of a typical schematic is shown in Fig. 8.35(b). With low- Z , plastic scintillators, the gamma interactions are primarily due to Compton scattering, and the maximum energy deposited in the scintillator is 340

keV for interactions of 511 keV gammas, and about 1 MeV for the ^{60}Co gammas. For investigations based on ^{60}Co , an energy window spanning 0.95 to 1.05 MeV is typical for the start detector, thereby ensuring minimal jitter in this PMT. The same narrow energy window, but in this instance of variable mean energy setting, is applied to the stop channel. A timing distribution is measured at each energy setting, covering the range 100 to 1000 keV from which the $1/\sqrt{E}$ dependence can be verified. For slow scintillators of low energy, timing is determined almost entirely by the statistics of the scintillator. For a CFD, $(Q/R_m)^{1/2}$ is constant, and the timing resolution should vary as τ_s/\sqrt{E} . However, at higher energies, transit-time jitter in the PMT dictates timing performance, as the contribution from scintillator statistics becomes increasingly insignificant with increasing energy. Consequently, the slope of the experimental data ultimately flattens with increasing energy.

Measurements of Moszyński (1993a), based on the performance with ^{22}Na and ^{60}Co , also show this $1/\sqrt{E}$ dependence for NaI(Tl), even at these relatively high energies. A measured resolution as fine as 156 ps for a scintillator of time constant 250 ns may appear surprising at first, but remember that the initial emission rate of NaI(Tl) is $R/\tau_s = 40 \text{ pe/ns MeV}^{-1}$, after the first 10 ns or so, and the statistics portend good timing; even more so for NE111, for which $R/\tau_s \sim 1500 \text{ pe/ns MeV}^{-1}$ and, when used with an XP2020UR-M, for example, the jitter is only 23 ps. The relationship between the dispersion and energy for NE111, which is very fast at 1.36 ns, follows a $1/(E)^{0.3}$ power law, indicating that the overall timing performance is dominated by the time jitter of the PMT and not by the scintillator statistics.

8.10 Timing summary

The laws governing electron motion in an electrical field underlie the timing performance of PMTs. The major contribution to time dispersion arises from the different paths taken between the photocathode and the first dynode. Electrons emitted with zero velocity and normal to the surface of the photocathode (plano-concave window assumed) are isochronous in their transit to d_1 . However, a spread in initial energy and an angular dependence in emission results in variable photoelectron transit time. These effects are mitigated in designs that provide a strong extraction field at the cathode. This translates into the use of a high potential difference between the cathode and the first dynode. Given the spread in initial energies with an assumed angular distribution, it is possible to predict the spread in transit times.

The role of Newton's laws of motion in the evolution of PMTs can be traced by the development of fast PMTs, notably by Philips starting in the 1960s with the well-known XP2020 and XP2040 series. The use of accelerating rings added to the front end, and single wires located between dynodes, indicate the importance

of boosting field gradients wherever feasible. More recently, major advances in speed of response are seen in the proximity focussed PMTs from Hamamatsu: a combination of miniaturization and high field strengths have produced devices with dispersion quoted in picoseconds.

In circuit analysis, given the response of a circuit to a δ function, the response to any input function is determinable. A δ -function input can be provided by single-photoelectron excitation, producing the signature of the anode output. This is described in terms of a pulse with finite rise time followed by a slower fall time, but for general simulation purposes it is sufficient to assume a simple exponential function, ignoring the finite rise time. Based on this, we are able to predict the shape of the anode output initiated by scintillators, for example, with the predictions bearing a close resemblance to actual oscilloscope traces. The surprising result for many is the degree of superimposed noise on pulses containing relatively large numbers of photoelectrons. This should prompt a question, especially with events of low photoelectron number: how will the electronics handle such a highly structured event? One measure of the degree of understanding of PMT operation is the degree of correlation between Monte Carlo simulations and actual recorded traces.

The occurrence of an event is signalled by the appearance of a pulse at the anode of a PMT, but the basis for determining the time at which the event occurred is complicated by the need to accommodate a wide range in pulse sizes: in basic electronic circuits, the large amplitude pulses will appear to arrive earlier than the smaller ones. This anomaly can be overcome by using the ingenious scheme of CF discrimination. Discriminators based on this technique are invariably preferred over those using other methods such as threshold detection and zero crossing. The performance of commercially available timing systems has reached the stage where the transit time of light within the bulk of the scintillator or Cerenkov converter may contribute more to the overall dispersion than do the PMT and the electronics.

Morton in the 1950s, and Gatti and co-workers some years later, formulated the general theory of time dispersion in terms of the contributions from the cathode to the first dynode, between dynodes, and from the last dynode to the anode. Poisson statistics were assumed for this purpose, from which an important generalization emerges: the fewer the number of stages, the smaller is the dispersion. The RCA 5-stage C31024, based on this principle, is considerably faster than the 12-stage, parent-type 8850. Specifying the timing dispersion for single photons provides a means for ranking PMTs but to predict and quantify the performance of a PMT–scintillator combination requires analysis of their combined statistics. An unexpected prediction from the statistical treatment of Post and Schiff (1950) is that best timing results from measuring the arrival of the first photoelectron. However, such a low-level signal cannot be measured at the photocathode, making it necessary to derive a timing distribution using the signal from the last dynode or anode. This results in a composite distribution with contributions from the photon statistics and those of the PMT. A Monte Carlo simulation, such as that in Fig. 8.39, indicates that the best timing favours a multi-photoelectron threshold. This,

fortuitously, is in accord with the preferred threshold requirement of CF discrimination: that is, an effective threshold part way up the leading edge of the signal.

Intuitively, one might expect that matching the timing parameters of the chosen scintillator to those of a selected PMT would give satisfactory and cost-effective performance. This is not so, as it is found with slow scintillators, such as the inorganic ones, that timing performance benefits from the use of fast PMTs—that is, those characterized by sub-nanosecond jitter. Moszyński quotes a figure of 160 ps for NaI(Tl) at the energy of ^{60}Co (Moszyński 1992; Moszyński *et al.* 2006). The reason behind the achievement of such performance with relatively slow inorganic scintillators is that timing depends on the initial rate of charge delivery, on the rise time of the scintillator output, and on the number of photoelectrons. For scintillators, with fast rise time, timing generally improves as R/τ_s .

REFERENCES

- Apker, L., Taft, E. A., and Dickey, J. (1953). Electron scattering and the photo-emission from caesium antimonide. *J. Opt. Soc. Am.*, **43**, 78–80.
- Bazarov, I. V. *et al.* (2008). Thermal emittance and response time measurements of negative electron affinity photocathodes. *J. Appl. Phys.*, **103**, 054901–8.
- Beck, G. (1976). Operation of a IP28 photomultiplier with subnanosecond response time. *Rev. Sci. Instr.*, **47**, 537–41.
- Bengtson, B. and Moszyński, M. (1982). Timing improved by the use of dynode signals studied with different scintillators and photomultipliers. *Nucl. Instr. and Meth.*, **204**, 129–40.
- Braunsfurth, J. and Körner, H. J. (1965). Resolving time properties of NaI scintillators. *Nucl. Instr. and Meth.*, **34**, 202–12.
- Collazuol, G. *et al.* (2007). Single photon timing resolution and detection efficiency of the IRST silicon photomultipliers. *Nucl. Instr. and Meth. in Phys. Res. A*, **581**, 461–4.
- de la Barre, F. (1973). Influence de l'éclairement sur la résolution temporelle des photomultiplicateurs. *L'onde électr.*, **53**, 133–40.
- de Vries, J. and Kelling, F. E. T. (1988). Fast timing with photomultiplier dynode pulses. *IEEE Trans. NS*, **35**, 392–5.
- Derenzio, S. E. *et al.* (2000). Measurements of the intrinsic rise times of common inorganic scintillators. *IEEE Trans. NS*, **47**, 860–4.
- Electron Tubes Ltd. (1996). *Photomultipliers and Accessories*, PMC/96. Electron Tubes Ltd., Ruislip.
- Euling, R. (1964). Statistics of time in photomultiplication. *J. App. Phys.*, **35**, 1391–403.
- Feller, W. (1957). *An Introduction to Probability Theory and its Applications*, Volume 1. New York, John Wiley and Sons, Inc., New York.
- Gatti, E. and Svelto, V. (1964). Revised theory of time resolution in scintillation counters. *Nucl. Instr. and Meth.*, **30**, 213–23.

- Gatti, E. and Svelto, V. (1966). Review of theories and experiments of resolving time with scintillation counters. *Nucl. Instr. and Meth.*, **43**, 248–68.
- Gedcke, D. A. and McDonald, W. J. (1968). Design of the constant fraction of pulse height trigger for optimum time resolution. *Nucl. Instr. and Meth.*, **58**, 253–60.
- Gedcke, D. A. and Williams, C. W. (1968). *Ortec Information Sheet, High Resolution Time Spectroscopy: I. Scintillation Detectors*. Ortec, Oak Ridge, TN.
- Hamamatsu Photonics K K. (2002). *Photomultiplier Tubes*. Hamamatsu Photonics K K, Hamamatsu City.
- Hungerford, G. and Birch D. J. S. (1996). Single-photon timing detectors for fluorescence lifetime spectroscopy. *Meas. Sci. Tech.*, **7**, 121–35.
- Hussain, K. A. and Butt, D. K. (1980). Time resolution studies of fast photomultipliers. *Nucl. Instr. and Meth. in Phys. Res. A*, **173**, 471–5.
- Kawai, Y. *et al.* (2007). Large aperture hybrid photo-detector. *Nucl. Instr. and Meth. in Phys. Res. A*, **579**, 42–5.
- Kerns, C. R. (1967). Photomultiplier single electron time spread measurements. *IEEE Trans. NS*, **14**, 449–54.
- Lakes, R. S. and Poultney, S. K. (1970). Subnanosecond single-photoelectron timing resolution of the RCA C31000E photomultiplier using the Ortec constant fraction discriminators and the Monsanto light emitting diodes. *Rev. Sci. Instr.*, **41**, 1889–90.
- Lerche, R. A. and Phillion, D. W. (1991). Rise time of BC-422 plastic scintillators < 20 ps. *IEEE Conference Record of the 1991 Nuclear Science Symposium and Medical Imaging Conference, 2–9 Nov. 1991*. IEEE, Piscataway, NJ, 167–70.
- Leskovar, B. and Lo. C. C. (1972). Performance studies of photomultipliers having dynodes with GaP(Cs) secondary emitting surface. *IEEE Trans. NS*, **19**, 50–62.
- Leskovar, B. and Lo. C. C. (1975). Single photoelectron time spread measurements of fast photomultipliers. *Nucl. Instr. and Meth.*, **123**, 145–60.
- Leskovar, B. and Lo. C. C. (1978). Time resolution performance studies of contemporary high speed photomultipliers. *IEEE Trans. NS*, **25**, 582–90.
- Metz, A. J. (1967). Simple tunnel diode circuit for accurate zero crossing timing. *Rev. Sci. Instr.*, **38**, 1445–9.
- Moszyński, M. (1992). Timing properties of Philips XP2020UR photomultiplier. *Nucl. Instr. and Meth. in Phys. Res. A*, **324**, 269–75.
- Moszyński, M. (1993a). Prospects for new photomultipliers. *Nucl. Instr. and Meth. in Phys. Res. A*, **337**, 154–64.
- Moszyński, M. (1993b). Timing properties of Philips XP2020UR photomultiplier. *Nucl. Instr. and Meth. in Phys. Res. A*, **334**, 269–75.
- Moszyński, M. and Vacher. J. (1977). Influence of the incident light wavelength on time jitter of fast photomultipliers. *Nucl. Instr. and Meth.*, **141**, 319–23.
- Moszyński, M. *et al.* (2004). New fast photomultipliers with a screening grid at the anode. *IEEE Trans. NS*, **51**, 1701–6.
- Moszyński, M. *et al.* (2006). New Photonis XP20D0 photomultiplier for fast timing in nuclear medicine. *Nucl. Instr. and Meth. in Phys. Res. A*, **567**, 31–5.

- Nathan, R. and Mee, C. H. (1970). Energy distribution of photoelectrons from K_2SbCs photocathode. *Phys. Stat. Sol. (a)*, **2**, 67–72.
- Nicholson, P. W. (1974). *Nuclear Electronics*. John Wiley and Sons, Inc., New York.
- Nutt, R. *et al.* (1970). A comparison of constant fraction and leading edge timing with NaI(Tl) scintillators. *IEEE Trans. NS*, **17**, 299–306.
- Philips. (1994). *Photomultiplier Tubes*, Pub. D PMT AB. Philips Photonics, Brie, France.
- Photek. (1994). *Ultra-Fast Photomultipliers* (datasheet). Photek, East Sussex.
- Piétri, G. (1973). ‘Progress in PMT for high energy physics instrumentation’. In *1973 International Conference on Instrumentation for High Energy Physics*, ed. S. Stipcich, p. 586.
- Piétri, G. (1981). *Photomultiplicateurs*, Pub. RTC 5482–07–07/1981. R.T.C. La Radiotechnique-Compelec, Cedex.
- Post, R. F. and Schiff, L. I. (1950). Statistical limitations on the resolving time of a scintillation counter. *Phys. Rev.*, **80**, 1113.
- Poultney, S. K. (1972). Single photon detection and timing: Experiments and techniques. *Advan. Electron. and Electron Phys.*, **31**, 39–117.
- Rozsa, C. M., Menge, P. R., and Mayhue, M. R. (2009). *BrillLanCe™ Scintillators Performance Summary (Revision: January 2009)*. Scintillation Products Technical Note. Saint-Gobain, Paris.
- Sipp, B., Miehe, J. A., and Lopez-Delgado, R. (1976). Wavelength dependence of the time resolution of high speed photomultipliers used in single-photon timing experiments. *Optics Comm.*, **16**, 202–04.
- Spangenberg, K. R., (1948). *Vacuum Tubes*, McGraw-Hill, New York.
- Spicer, W. E., (1972). ‘Photoelectric emission’. In *Optical Properties of Solids*, ed. F. Abeles, Elsevier, Amsterdam, 755–858.
- Tahir, K. *et al.* (1991). ‘Computer modelling of the electron optics of photomultipliers.’ In *Proceedings of the 10th Symposium on Photoelectronic Image Devices*, pp. 429–36.
- Weber, M. J., Derenzo, S. E. and Moses, W. W. (2000). Measurement of ultrafast scintillation rise times: Evidence of energy transfer mechanisms. *J. Lumin.*, **87–9**, 830–2.

9

Linear performance

9.1 Introduction

There is no generally accepted formulation for expressing linearity, let alone one for measuring it. Some methods for quantifying non-linearity in DC detection may apply to pulsed signals. In this section, we explore a range of ways in which this parameter may be measured; it is important to choose a method that mimics the intended application. Analogue measurements refer to anode current—also termed DC detection. The concern with pulsed signals is linearity of charge output, for which there are two ways to express this non-linearity: differential and integral. This categorization appeared in the 1950s in connection with stacked discriminators, the forerunner of multichannel pulse height analysers of the Wilkinson type. A clear account of the early treatment of linearity in pulse height encoding is given by Nicholson (1974), and a current treatment of non-linearity in ADCs may be found in the application notes issued by manufacturers of such products (see e.g. Texas Instruments 1995).

Transducers are detectors for measuring physical entities, with the result expressed in a different variable or form. PMTs are such devices, since they convert a light signal into an electrical analogue. In common with all transducers, there is a restricted extent over which a PMT can be expected to function satisfactorily: this is known as the dynamic range, defined as the ratio of the maximum to the minimum signal, subject to meeting a given performance specification. Operation outside these limits invariably has undesirable consequences: results may be in error, but more seriously, a PMT may become damaged. Note the International Electrotechnical Commission (2010) does not define dynamic range and its dependency on setting upper and lower performance limits, perhaps recognizing that this parameter should form the basis of a specification between a user and a manufacturer, and devised for a particular application. Attention in this chapter is focussed on a single performance parameter: the degree of proportionality between an incident light signal and its resulting electrical output. This is referred to as linearity, which usually applies to the anode signal, but it is also appropriate for a dynode signal. The quality of performance depends on the details of the application, the suitability of the chosen PMT, and the method of operation: that is, whether it is photon counting, analogue (DC anode current), or pulsed (charge). Non-uniformities in photocathode QE, resistivity, and collection by the multiplier are also sources of non-linearity but they are discussed in 2.10. Note that a linearity specification may refer to a complete system, consisting of a PMT and a preamplifier, for example.

Discussion of the terminology is appropriate at this juncture to avoid confusion in the sections that follow. When the output signal is a slowly varying current, with changes occurring over relatively long time periods—for example, milliseconds or seconds—the operation is described as analogue; the mean anode current is the key abscissa parameter. Alternatively, where fast pulses of light are under

consideration—for example, events of nanosecond or microsecond duration—the output takes a similar form and appropriate variables are peak anode current, or the charge in each anode pulse. PMT pulses are seldom symmetrical in rise and fall time but the peak pulse height is proportional to the total charge, provided linear operation applies: this facilitates the convenience of making measurements on an oscilloscope. Non-linear effects are associated with both mean current levels and with pulsed currents. It is therefore important when determining charge linearity to exclude any contribution originating from changing the event rate. Although the effects of the latter are equally important, quantifying this form of non-linearity involves a different procedure. Charge and pulse height non-linearity must be measured at a low event rates, corresponding to negligible mean anode current.

9.1.1 Modes of operation

PMT applications fall into two categories: the magnitude of the signal, in pulsed applications is quantified either with reference to the area (charge) of the pulse, or to its height quoted in millamps. Signals range in size from single-photoelectron initiated to pulses containing up to about 10^5 pe. In the second category, signal information resides in the mean anode current. Non-linearity manifests itself in different ways in the two modes of operation. Single photoelectrons, even after amplification by a gain of 10^8 , are relatively small in magnitude and pulse height linearity is preserved for individual pulses. At the opposite extreme, pulsed anode signals, initiated by inputs containing greater than about $\sim 10^3$ pe, frequently show non-linearity, particularly if the pulses are fast. Achieving a wide dynamic range of operation is challenging with ultra-fast signals, from a plastic scintillator, for example, compared with the relatively slow pulses from NaI(Tl). In the former case, the half-width of the pulse is of the same order as one initiated by a single electron (a few nanoseconds), while for sodium iodide it is ~ 250 ns, for example. Achieving a high degree of linearity in gamma ray spectroscopy using a NaI(Tl) crystal is therefore not as challenging as it is for the faster scintillators. This is illustrated in Fig. 9.1, which shows two pulses of the same charge but with different peak currents. The linearity and hence the dynamic range will always be superior with pulses of Type (b).

An obvious indication of non-linearity is presented in Fig. 9.2, where the typical sign is a flattening of the peak and an extension to the tail of the pulse, as in the profile labelled (b). This form of non-linearity is attributed to space charge saturation occurring between pairs of dynodes. It is described by the Langmuir–Child three-halves power law: that is, the saturation varies as $(V(d_{n+1}) - V(d_n))^{3/2}$. Such non-linear behaviour is tolerable in photon and scintillation counting, since pulses merely have to exceed an imposed low-level threshold to register. The only requirement, therefore, is that the PMT should be linear in the vicinity of the set threshold. However, non-linearity ultimately sets in at high rates because of loss in counts through pulse overlap—a consequence of increased dead time.

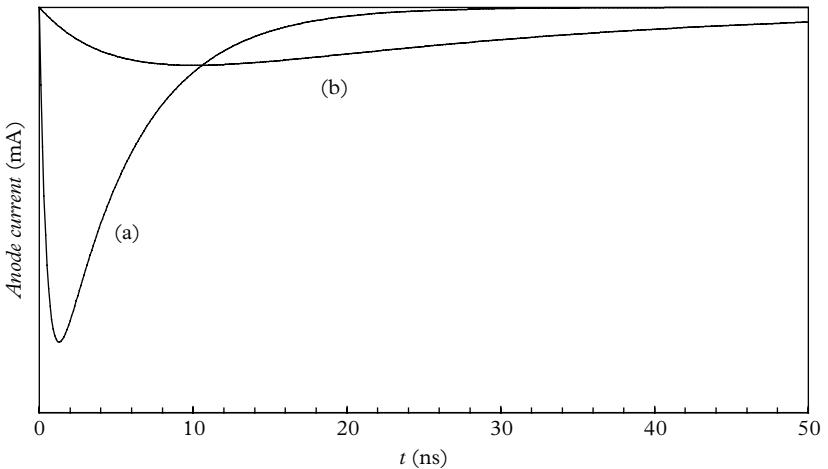


Fig. 9.1. Representations of multi-electron anode signals for (a) a fast Cerenkov like pulse and (b) typical of one produced by a scintillator with a relatively long time constant. The area under each pulse is the same but the different peak heights should be noted.

9

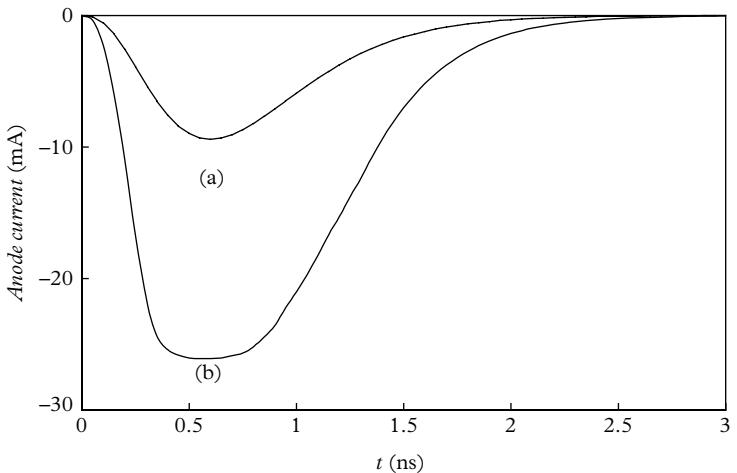


Fig. 9.2. How non-linearity in gain distorts the shape of an output pulse. The profile (a) represents linear operation, while (b) shows the effects of overload on big pulses. Non-linearity is a maximum at the peak of the pulse.

9.1.2 Rate effect

There is an inescapable phenomenon in both modes of operation that is associated with an all-resistor voltage divider. Drawing anode current always produces a redistribution of inter-dynode voltages and consequently an increase in gain. It is important to recognize that this type of non-linearity, known as rate effect, is not intrinsic to the PMT. It stems from changing bias conditions in the voltage divider, as the mean anode current varies in response to a change in the intensity of the light signal. The phenomenon applies to analogue and pulsed applications and, while there are means to limit this type of non-linearity, it is always present to some degree where a resistive divider is used. Its nature is verified by an analytical treatment in 13.3. Such a contribution to non-linearity is particularly troublesome at mean anode currents in excess of $\sim 10 \mu\text{A}$ but performance can be greatly improved by replacing a resistive voltage divider with an active one, following the discussion in 13.6.

9.1.3 Causes of non-linearity

It is important to know where the boundaries to linear performance lie: for low-level signals, PMT parameters such as leakage and dark current are relevant indicators; at high light levels, space charge build-up, power dissipation, and voltage redistribution in the divider ultimately set an upper limit to satisfactory performance. Mesh dynodes and linear focussed dynodes provide the best linearity at high signal levels, while box-and-grid and Venetian blind designs are more restricted in their dynamic range. Extension to linear performance in pulsed applications can be achieved by using a tapered voltage divider to defer the onset of space charge saturation. Such a divider allocates higher inter-dynode voltages to the backend of a PMT, leading to the performance shown in Fig. 9.3.

9

9.1.4 Dynode materials

There is a source of anode current non-linearity intrinsic to the material of the dynodes but unrelated to the behaviour of the divider. Historically, certain PMTs, particularly those with BeCu dynodes, suffered in this respect by showing an increase in gain with mean anode current (see e.g. De Vincenzi *et al.* 1984; ET Enterprises 2001, Fig. 27). For historical reasons, this non-linearity was termed ‘rate effect’ because it first became apparent in testing PMTs for NaI(Tl) applications, such as gamma cameras. One of the tests involved observing the peak shift of ^{137}Cs following a ten times increase in the source rate.

It is well known that PMTs that include BeCu dynodes have higher peak current capability than their SbCs counterparts. However, tests by the author show little difference in pulse height linearity between the two dynode types when their respective PMTs are tested at the same overall voltage and at the same anode signal level. Note that different light levels apply in the two cases: the PMT with the BeCu dynodes requires a higher signal than the PMT with SbCs dynodes. Photocathode linearity must therefore be checked.

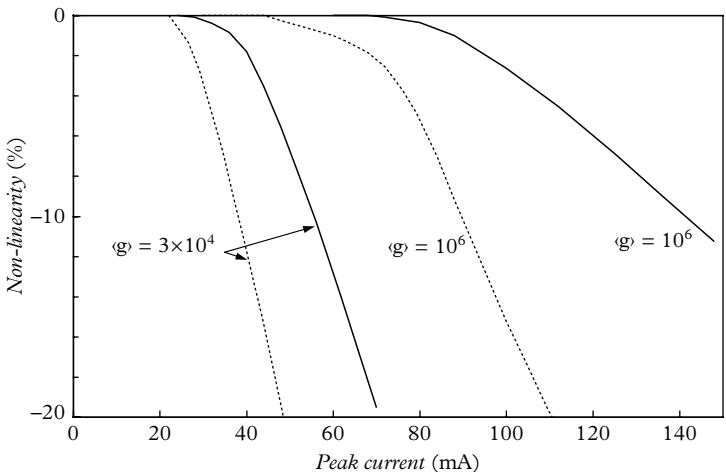


Fig. 9.3. Investigation of the pulsed linearity of a 9266B PMT in a tapered divider ($5R$, $R, \dots, R, 1.5R, 2R, 2.5R, 3R$) for pulses of 100 ns, fwhm. The dashed curves refer to peak pulse height linearity, and the solid curves to charge mode of operation. Non-linearity is determined using a pulsed version of (9.1), described in 9.2.4. The dependence of non-linearity on gain, $\langle g \rangle$, should be noted.

9.2 Dynamic range

As mentioned in 9.1, the term ‘analogue mode of operation’ refers to measurements made in terms of anode current. There are two considerations regarding linearity. First, where signals are single photon, as in photon counting, the anode current will consist of a stream of charge pulses, each initiated by a single photo-electron, and all producing approximately the same anode charge. These pulses are small in charge content, even at high PMT gain, and there is no associated non-linearity in either pulse height or pulse area. Even at high count rates, where pulses may overlap, anode current remains linear with respect to light flux, given an ideal voltage divider. Fatigue and PMT gain shift ultimately constrain linear performance to anode currents below $\sim 10 \mu\text{A}$, where all-resistor dividers are used, but active voltage dividers can provide linear performance up to $300 \mu\text{A}$ (see 9.7, Fig. 9.16). Second, where detection is analogue, but light signals are multi-photon, linearity will depend on the signal rate and on any pulse height non-linearity of the PMT, such as that given in Fig. 9.2.

9.2.1 Achievable dynamic range: Analogue

We first consider analogue detection, showing that it may be possible to operate over a dynamic range of the order of 10^7 , but what is actually achievable depends

critically on the type of PMT and on the quality of performance required by the user. A reasonable starting point for discussion is to set the lower limit for the dynamic range equal to the dark current, and to set the upper limit at the maximum anode current quoted by the manufacturer: this leads to a dynamic range of 10^5 by assuming typical manufacturer's figures of 1 nA and 100 μ A, for the two respective limits. Although these figures may appear subjective, the author is unaware of an alternative approach for defining this parameter. The dynamic range can be extended by selecting a PMT with low dark current, which may be further reduced on cooling. In which case we might gain a reduction in the lower limit to 0.1 nA, for example, leading to an extension of the dynamic range to 10^6 . The upper limit to the dynamic range may be increased to say 300 μ A by using an active divider but for short-lived excursions only (seconds rather than minutes) if fatigue is to be avoided. The dynamic range in this example is now 3×10^6 , based on a lower limit equal to the reduced dark current; arguably, given sufficient measurement time, this could be set to one-tenth of the dark current, yielding a dynamic range of 3×10^7 .

9.2.2 Achievable dynamic range: Pulse mode

The dynamic range that can be achieved under pulsed conditions is determined primarily by the onset of inter-dynode space charge saturation. The phenomenon is best understood in terms of the movement of charge between dynodes: in other words, the controlling influences are the transient currents flowing between dynodes, and therefore the effect is greatest at the ultimate and penultimate dynodes, where peak signal currents are highest. From a practical point of view, the higher the peak anode current, the more likely is the occurrence of non-linear amplification. This is demonstrated by comparing the shape of signals from a Cerenkov detector with those from a scintillator. Light output is effectively instantaneous for a Cerenkov signal, and the shape of the anode output should closely mimic that of a scaled-up single-photoelectron event. An idealized fast pulse is shown in Fig. 9.1(a), together with a broader example, shown in Fig. 9.1(b), which is generated from a scintillator of longer time constant. Both signals are normalized in charge, but the signal in the scintillator pulse is more dispersed over time, leading to a peak current of only about one-fifth that of the Cerenkov pulse. This has the effect of generating an extension to the dynamic range. A linear focussed PMT with a tapered divider can deliver a peak pulse height of ~ 200 mA and, assuming at the other extreme of the dynamic range that 1 mV is just discernible on an oscilloscope, a dynamic range of $\sim 10^4$ is possibly achievable. The only sensible way to increase the dynamic range in this example is to include a fast amplifier.

Figure 9.2 shows the effect of non-linearity as a flattening of the peak at high light levels. There are two sources of non-linearity in spectroscopy—one is caused by pulse height compression, and the other by a variation in signal rate. Although the two effects are contrary in influence, their contributions cannot be balanced. The way in which non-linearity manifests itself with pulsed light is more complicated

than for DC sources, but we can simplify the analysis by first treating those applications in which pulses are counted, with no requirement to size them. In photon counting, for example, non-linearity between light flux and count rate is covered in 14.9 under the heading of dead time, and this aspect of PMT non-linear performance need not to be considered here. Variability in pulse heights, and even saturation in the larger signal amplitudes, does not necessarily lead to non-linearity in single-photon detection. In this mode of operation, pulses only need exceed a set threshold to be counted. The basis of operation in certain nuclear radiation probes (e.g. ${}^3\text{H}$ and ${}^{14}\text{C}$ monitors) is that of counting multi-photoelectron pulses and, as with photon counting, the primary source of non-linearity stems from dead-time losses. As has been previously discussed, there is always a gain shift with count rate in a PMT when used in an all-resistor divider, but one of the merits of threshold detection is the near independence of count rate with changes in gain, discussed in 7.6.4.

By comparison, in spectroscopy, non-linearity in gain, as a function of pulse height, is central to the specification. It is important to appreciate that peak anode current non-linearity and anode charge non-linearity are related phenomena. Although current and charge are always linked in the sense that current is the differential of charge with respect to time, non-linearity affects these two parameters in different ways. This can be illustrated by reference to the pulsed signals shown in Fig. 9.2. A typical PMT output pulse is represented by (a) when the gain of the PMT is linear with respect to pulse height, and by (b) in the presence of non-linearity. Note this PMT operates linearly for pulsed anode currents below -10 mA ; thereafter, there is a loss of area until the tail of the signal passes 10 mA . The difference in magnitude between the two parameters always applies, regardless of pulse shape, with the magnitude of peak current non-linearity always exceeding that of charge non-linearity. Telltale signs of severe non-linearity are a flattening of the peak, and a time extension in the signal tail.

Mean anode current non-linearity has been identified as one of the limiting factors in the achievement of wide dynamic range. A significant cause of non-linear amplification in analogue mode stems from malfunctioning of the voltage divider—normally, an all-resistor type. It has been shown in 13.3 that a voltage divider consisting of a series-connected array of resistors cannot maintain fixed bias voltages, thus leading to unstable gain, in the presence of a variable mean anode current, I_a . Generally, acceptable performance, characterized by gain stability of $\sim 1\%$, can be achieved with such a voltage divider by setting the standing bias current, I_{D0} , to 100 times the anticipated maximum mean anode current. This empirical rule is long established, and supported by theoretical analysis in 13.3. These restrictions obviously impose a penalty on achieving wide dynamic range since I_{D0} (the divider standing current in the absence of signal) should not be set higher than $\sim 1\text{ mA}$, if power dissipation in the divider is to be avoided. The ‘100 times rule’ implies an upper limit to anode current of $\sim 10\text{ }\mu\text{A}$, as previously mentioned.

Where pulses are concerned, linear operation can be extended to signals with a peak current of 100 mA , as shown in Fig. 9.3. However, with a resistive divider this

performance can only be achieved at low count rates, if contributions to non-linearity from the rate effect are to be avoided. The demands of a high count rate application can be met, but only with an active divider. The major source of non-linearity resides in the back end of the PMT, and improved pulse performance, up to peak anode currents approaching 200 mA, can be achieved by tapering. However, it is shown in 13.6 that tapering accentuates gain shift with mean anode current, leading to a reduction in dynamic range. One way to avoid this problem is to stabilize the bias voltages on the last two or three dynodes by using a combination of zener diodes to achieve the required degree of taper. A better solution is to include MOSFETs in source-follower mode, requiring as little as 20 µA quiescent supply current, I_{D0} , with potential to provide linear performance up to mean anode currents of three hundred microamps. This is feasible while simultaneously achieving linear amplification for peak currents in excess of 100 mA (see 13.6).

9.2.3 Onset of DC non-linearity

Detailed methods for quantifying non-linear performance are presented in 9.2.4, but the immediate aim is to find ways for signalling its onset. The method for DC linearity involves two independent and random LED light sources, A and B, mounted within the same enclosure and powered by individual low voltage supplies; the light sources are viewed by a PMT operated at fixed HV. The intensity of each light source is set to give approximately the same anode current. With the two anode signals referred to as I_A and I_B , respectively, a PMT produces an output of I_{A+B} when the two sources are both switched on. While the PMT is operating linearly, it is clear that $I_{A+B} = I_A + I_B$. The onset of non-linearity is evident when the coincident sum current, I_{A+B} , ceases to equate to the arithmetic sum, $I_A + I_B$. Measurements must start at low light levels, where there is negligible non-linearity—established by verifying that $I_{A+B} = I_A + I_B$, in accordance with (9.1). A set of measurements is recorded by incrementing both I_A and I_B , more or less equally. The procedure outlined here may be followed by measuring the voltage developed across an anode load resistor, if such a component is included in the application; or, current may be measured directly. Finer steps in light-level adjustment are suggested once non-linearity first becomes evident. Integral non-linearity is calculated from

$$\epsilon = \frac{I_{A+B} - (I_A + I_B)}{I_A + I_B}. \quad (9.1)$$

The dual LED method described above is valid for determining small deviations from anode current linearity because it noticeably underestimates the true non-linearity once it exceeds ~5 %. The reason for this is that light levels are progressively over incremented once non-linearity is evident, leading to incorrect location of ϵ values—they are set too far to the right on the abscissa. This occurrence has no effect while the PMT is operating linearly. Hence, the method may be used as a

Table 9.1. Non-linearity determination, ϵ , using two light sources. Current entries are in microamps.

I_A	I_B	I_{A+B}	$I_A + I_B$	ϵ (%)
5.01	4.98	10.03	9.99	0.40
9.99	10.04	20.24	20.03	1.05
15.06	15.01	30.55	30.07	1.60
20.06	20.12	41.05	40.19	2.14
25.11	25.06	51.54	50.17	2.73
30.11	29.99	62.15	60.10	3.41
35.20	35.04	73.15	70.24	4.14
40.14	40.00	84.00	80.14	4.82
45.27	45.00	95.26	90.27	5.53
50.03	50.40	106.74	100.43	6.28

specification based on a maximum permitted non-linearity figure; this must be small to avoid accumulating errors and is typically within the range of 1 to 2%. The merit of the method lies in its simplicity, and ease of measurement, making it a useful way of ranking or comparing PMT performance. Dark current subtraction is necessary for anode signal currents comparable to the dark current, which is included twice in $I_A + I_B$ but only once in I_{A+B} . Table 9.1 shows a set of results taken with two LEDs in this way, using a conventional all-resistor divider.

9.2.4 Onset of pulsed non-linearity

The procedure described in 9.2.3 can be applied to pulsed light sources, such as LEDs. There are two aspects to consider: pulse height linearity and charge linearity. Pulse height linearity may be determined with an oscilloscope or peak-sensing electronics. Charge linearity refers to the total charge within a pulse, quantified by integrating the anode current. This may be accomplished by a passive $R \parallel C$ network, with a charge-sensitive amplifier–MCA combination, or by using digital-signal-processing instrumentation. The two aspects of non-linearity are not the same, with charge linearity always providing a wider dynamic range than peak linearity does. The method for pulses shares the same basis as that of 9.2.3 and with the same shortcoming. Two pulsed light sources, P_A and P_B , mounted within the same enclosure and powered ideally by independent pulse generators, are viewed by a PMT under test. The shape of each pulse must be common, and compatible with the application. For example, an exponential decay of time constant 250 ns is representative of NaI(Tl) applications. The intensity of

each light source is set to give approximately the same pulse height on an oscilloscope. Equation (9.1) is the basis for the determination of ϵ , with I_A and I_B replaced by pulse heights P_A and P_B , respectively. The method involves noting the peak heights of P_A and P_B taken individually, followed by measurement of the pulse height when both are active and coincident. It is important in this and in other pulsed methods to avoid a contribution to non-linearity from the effects of changing mean anode current, as discussed in 9.1.2. Naturally, as pulse heights are increased so is the mean anode current but the present requirement is to determine non-linearity in pulsed mode alone, without incurring a contribution from the effects of changing the mean anode current. The latter varies in accordance with pulse height but this contribution may be effectively eliminated by choosing low count rates and hence low mean anode currents.

An improved method, essentially free from rate effects, can be achieved by arranging that P_B , for example, be delayed with respect to P_A —a facility included in most pulse generators. Both light pulsers are permanently on, and the measured pulse height data consist of coincident and anti-coincident measurements taken in accordance with the timing schematic of Fig. 9.4. The mean current is essentially the same in both modes of operation, regardless of signal phasing. This is not strictly true once non-linearity sets in but reducing the signal frequency mitigates this second order effect. The repetition rate, $1/t_4$, may be adjusted between readings to maintain a constant DC level of anode current. As in the previous method, pulse heights can be read off an oscilloscope to yield the peak pulse height non-linearity.

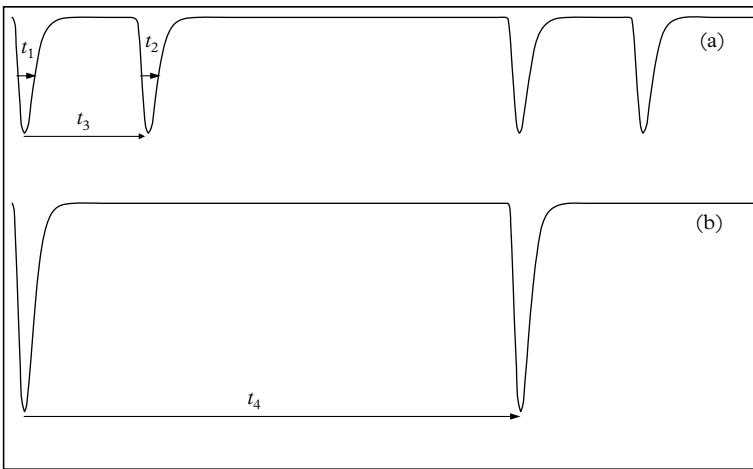


Fig. 9.4. Waveforms for pulsed LEDs, showing (a) out-of-phase and (b) in-phase operation, $t_3 = 0$. The signals t_1 and t_2 are set to the same width, and t_3 is adjusted so that anode pulses do not overlap. The time intervals t_4 and t_3 must be long compared with the parameters t_1 and t_2 . With linear PMT operation, the mean anode current is the same in both the (a) and the (b) configurations.

Table 9.2 The bootstrap method for determining integral non-linearity uses two light sources and overcomes the deficiency of the two-LED method described in 9.2.3.

I_A	I_B	I_{A+B}	φ	$(I_{A+B}-\varphi)/\varphi(\%)$
8	8	16	16	0.0
16	16	33	32	3.1
33	33	69	64	7.8
69	69	150	128	17.2

In addition to measuring peak non-linearity, an indication of charge linearity can also be investigated with reference to Fig. 9.4, by taking readings with a sufficiently damped multimeter. This arrangement is most convenient with negative HV, although adaption to positive polarity only requires a load resistor and a floating multimeter. It may appear puzzling that a DC method can be employed to measure pulsed non-linearity, until it is realized that what is actually being measured is the charge content of each pulse.

9.2.5 Bootstrap method

The flaw in the method described in 9.2.3 may be rectified by plotting φ readings at the correct abscissa location. This is illustrated in tabular form in Table 9.2, using representative readings. Two light sources are each set to 8 μ A, with the condition $I_A + I_B = I_{A+B} = 16 \mu$ A, thereby establishing and confirming linear operation at this current level; I_A and I_B are then individually set to the first I_{A+B} reading of 16 μ A, resulting in a sum of $I_{A+B} = 33 \mu$ A, representing 1 μ A of non-linearity. Then, I_A and I_B are both set to 33 μ A, generating a combined current of 69 μ A with a 3 μ A excess, and so on. It is clear that the entries in the φ column are increased by a factor of 2 in moving from one row to the next, resulting from doubling the light outputs at each step. The range of 0 to 128 μ A therefore contains only four readings, which can be supplemented by choosing a different starting value for I_A and I_B —say, at 10 μ A each. The stability of the system is paramount since setting errors accumulate. Note the bootstrap method applies equally to pulsed signals.

9.3 Theoretical considerations

The treatment so far concerns various forms of non-linearity, practical aspects of the effect, and its consequences. There are two ways for quantifying non-linearity: differential and integral. Numerical computations for both parameters can usually

be made from the same set of data but to enable this requires definitions and a mathematical treatment. A generalized approach for the measurement of integral linearity is followed, which is applicable to pulsed and analogue signals. Defining linearity in terms of summation we have, according to Aramu and Rucci (1967), that a PMT is linear if the anode signal, $I(\varphi)$, is proportional to the input light flux, φ , over a stipulated range in $I(\varphi)$. For two independent light sources, φ_1 and φ_2 , linearity requires that the output, when both signals are present, must equal the sum of the individual responses; thus,

$$I(\varphi_1 + \varphi_2) = I(\varphi_1) + I(\varphi_2). \quad (9.2)$$

For k independent light sources,

$$I(\varphi_1 + \varphi_2 + \dots + \varphi_k) = I(\varphi_1) + I(\varphi_2) + \dots + I(\varphi_k). \quad (9.3)$$

PMT performance is deemed non-linear whenever the two sides of these equations differ. Furthermore, if the left-hand side is greater than the right-hand side, the performance is said to be supre-linear; otherwise, the term non-linear applies. Note these equations are already familiar from (9.1) as a basis for measuring non-linearity.

Determination of PMT non-linearity based on (9.2) and (9.3) through experimentation may appear straightforward to the unwary, but in practice one must avoid the trap in setting the test light sources—this was touched on in 9.2.3. If the PMT is used to set the level of a test light source, any non-linearity in the anode signal will propagate. However, a set of k individual light signals can be used to measure non-linearity, according to these formulations, provided that individual φ_k settings fall within the region of linear operation of the PMT. It is convenient, although not essential, to set all φ_k to nominally the same intensity. If (9.2) is satisfied, then the PMT is linear, at least for light signals up to $\varphi_1 + \varphi_2$, and for corresponding anode signals up to $I(\varphi_1) + I(\varphi_2)$. These considerations form the basis of arguably the most direct method for measuring DC and pulsed linearity by using multiple light sources of adjustable intensity. The present author utilized ten adjustable light sources to obtain the results shown in Tables 9.3 and 9.4.

Integral non-linearity may be defined as

$$\Delta I/I = \Delta g/g. \quad (9.4)$$

where I represents both DC and pulsed operation; ΔI is the difference between an experimental point relative to a straight line, and $\Delta g/g$ represents a relative change in gain. Non-linearity specified in a differential form is a more stringent measure and it is defined as

$$\frac{\frac{dI}{I}}{\frac{d\varphi}{\varphi}} \text{ vs } I, \quad (9.5)$$

where φ relates to the input light level expressed as a current. Linearity applies when the dimensionless quantities dI/I and $d\varphi/\varphi$ are identical but, in the presence of non-linearity performance, dI/I deviates from $d\varphi/\varphi$.

Table 9.3 The outputs from ten light sources, listed in Column A, are set to nominally the same level.

A $d\varphi$	B φ	C I	D dI	E S	F $\varphi = \sum dI/S$	G K	H I/φ	I $1/K$
10.050	0.000	0.000	10.039	0.999	0.000	1.000	1.000	0.999
10.055	10.055	10.039	10.217	1.016	10.055	0.983	0.998	1.018
10.041	20.096	20.256	10.404	1.036	20.096	0.973	1.008	1.028
10.015	30.111	30.660	10.625	1.061	30.111	0.960	1.018	1.042
10.000	40.111	41.285	10.980	1.098	40.111	0.937	1.029	1.067
10.102	50.213	52.265	11.154	1.104	50.213	0.943	1.041	1.061
10.005	60.218	63.419	11.421	1.142	60.218	0.923	1.053	1.084
10.005	70.223	74.840	11.803	1.180	70.223	0.903	1.066	1.107
10.052	80.275	86.643	11.995	1.193	80.275	0.904	1.079	1.106
9.966	90.241	98.638	12.331	1.237	90.241	0.883	1.093	1.132
10.016	100.257	110.969						
			$C_n - C_{n-1}$	D/A	$(\sum D)/E$	C/(EF)	C/B	1/G

9

We transform measured quantities dI/I to $d\varphi/\varphi$ by assuming a correction function, $K(I)$, so that

$$\frac{d\varphi}{\varphi} = K(I) \frac{dI}{I}. \quad (9.6)$$

Here $K(I)$ is a function of I only and it can be determined from readings of dI and I ; $K(I)$ is unity for a detector operating linearly. The functional relationship between I and φ is illustrated in Fig. 9.5, together with the slope $S(I) = dI/d\varphi$, at any point on the curve. Now,

$$d\varphi = \frac{dI}{S(I)} \quad (9.7)$$

and, upon integration,

$$\varphi = \int_0^I \frac{dI}{S(I)}. \quad (9.8)$$

Substituting (9.7) and (9.8) into (9.6) generates

Table 9.4 Data taken with the step method. The step $d\varphi$ (Column B) was checked before each measurement of I – it does not vary by more than 1% over the measurement cycle. Note the data in Columns A, G, and J refer to φ , providing a useful check on computation.

A $\varphi = \sum d\varphi$	B set $d\varphi$	C I	D $I + dI$	E $(I + dI) - I$	F S	G $\sum dI/S = \varphi$	H $\sum I = I$	I $K = I/(S\varphi)$	J $\varphi = I/(SK)$
9.971	9.976	0.000	9.976	9.976	1.000	9.976	9.976	1.000	9.976
19.942	9.971	9.992	20.138	10.146	1.018	19.947	20.122	0.991	19.947
29.902	9.960	20.044	30.397	10.353	1.039	29.907	30.475	0.980	29.907
39.890	9.947	30.051	40.612	10.561	1.062	39.854	41.036	0.970	39.854
49.780	9.931	40.005	50.798	10.793	1.087	49.785	51.829	0.958	49.785
59.703	9.923	50.043	61.066	11.023	1.111	59.708	62.852	0.948	59.708
69.620	9.917	60.079	71.361	11.282	1.138	69.625	74.134	0.936	69.625
79.533	9.913	69.960	81.494	11.534	1.164	79.538	85.668	0.926	79.538
89.435	9.902	80.022	91.785	11.763	1.188	89.440	97.431	0.917	89.440
99.338	9.903	90.555	102.590	12.035	1.215	99.343	109.466	0.907	99.343
109.230	9.893	99.973	112.250	12.277	1.241	109.236	121.743	0.898	109.236
			D-C	E/B	(E/F)+G	E+H	H/(FG)	H/(FI)	

$$K(I) = (I/S(I)) / \int_0^I dI/S(I). \quad (9.9)$$

9

Note that $S(0)$ is unity for a quality PMT that is free from electrical leakage to the anode, with negligible dark current, and in a light tight enclosure. Rearranging, we have, using (9.8),

$$\varphi = \frac{I}{K(I)S(I)}. \quad (9.10)$$

The right-hand side of (9.10) is a function of I only, as required, and evaluating the integral leads to φ , the true input signal. Similarly, (9.8) also generates φ from I . For practical purposes, the integrals are replaced by sigmas, which is permissible provided the steps in I are sufficiently small, or the functional relationship between φ and I changes relatively slowly—or both. Equation (9.8) takes the form

$$\varphi = dI_0/S(I)_0 + dI_1/S(I)_1 + dI_2/S(I)_2 + \dots + dI_n/S(I)_n. \quad (9.11)$$

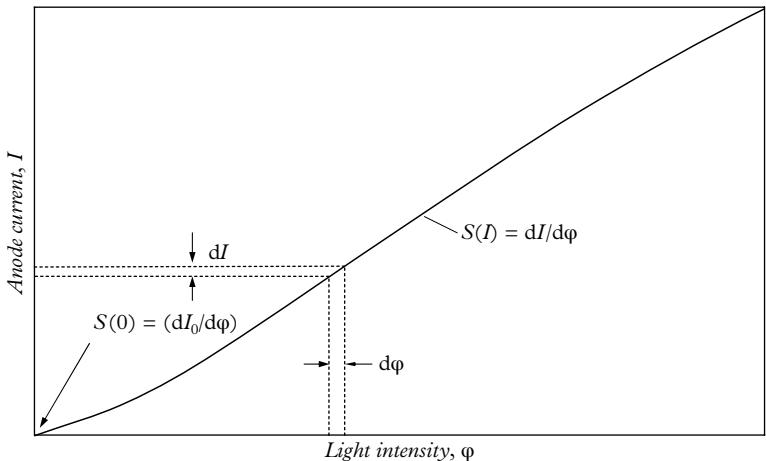


Fig. 9.5. A response curve illustrating the terminology in 9.3.

The essence of the method, as described by Fenster *et al.* (1973), is to uncover φ , by maintaining $d\varphi$ constant while I is varied; dI is the difference between adjacent readings of I .

9.4 Advanced linearity investigations

9

These methods involve the use of multiple independent light sources, which may be individually switched on or off. The method is suitable for pulsed or DC operation, and is illustrated in Table 9.3 for the analogue mode. Each light source is set to the same nominal brightness, using the PMT anode current for this purpose. Furthermore, the brightness of each source is set sufficiently low to locate I within the region of acceptable linear operation. This is checked by the successive application of just two light sources (A and B), requiring that the individual anode currents, I_A and I_B , sum to the coincident current I_{A+B} . It is clear in Fig. 9.6 that operation with acceptable linearity is achieved provided that $\varphi < 20 \mu\text{A}$; for this measurement series, φ is taken as $10 \mu\text{A}$.

In Table 9.3, Column A records the ten current settings, $d\varphi$, by individual adjustment of the LEDs. Column B shows the cumulative signal input, $\varphi = \sum d\varphi$, and Column C shows the value of the corresponding anode current; dI is derived from adjacent rows in Column C; S is the slope of the response curve, $dI/d\varphi$; and the summation in F leads to φ in Column B and serves as a consistency check. The correction factor $K = I/(S\varphi)$ is given in Column G. The coordinates (circles) in Fig. 9.6 refer to Columns B and C and clearly demonstrate supre-linear performance for anode currents in excess of $30 \mu\text{A}$. The good fit obtained with a quadratic

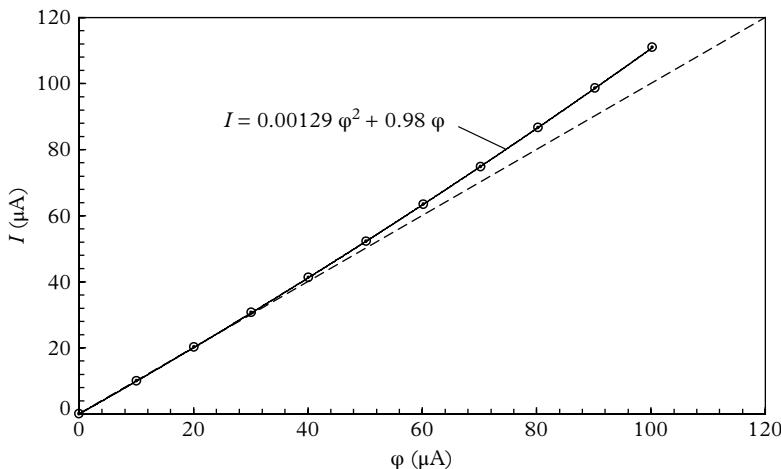


Fig. 9.6. A DC non-linearity curve derived using a set of ten LEDs. There is no discernible non-linearity for $\varphi < 20 \mu\text{A}$. Note that the quadratic relationship coincides very closely with the measured coordinates. The dashed line represents linear amplification.

is noteworthy. Such a relationship is ideal for devising a look-up table for correcting real data; the requirement for this purpose is also a quadratic but with the axes interchanged, as in Fig. 9.7.

9.4.1 Single-step method

The method just described has the disadvantage of requiring multiple light sources, all of which must be free from crosstalk. A method devised by Gatti and Piva (1953) to overcome the problems of differential non-linearity in pulse height encoding may be adapted for measuring non-linearity in DC and pulsed signals in PMTs; this method has the advantage of using only two light sources. In the application of this method, non-linearity is determined by noting the changing amplitude of a small fixed-light pulse when coincident with the principal one of variable height, as illustrated in Fig. 9.8 (this is sometimes referred to as the ‘piggyback’ method). In DC detection, a small incremental current of fixed amplitude, $d\varphi$, is added to an existing current I to give $I + dI$; the difference between $[(I + dI) - I]$ and $d\varphi$ leads to a measure of non-linearity. Once non-linearity sets in, the incremental anode current contribution dI continues to change as I is increased. The functional relationship between dI and I leads to the differential non-linearity dI/I .

Table 9.4 is a record of readings taken under the step method, with $d\varphi$ set nominally to $10 \mu\text{A}$. The stability of this light source was monitored at each setting of I and recorded in Column B; I versus the φ coordinates C and G is included in Fig. 9.9 for direct comparison with the method based on ten light sources. Here

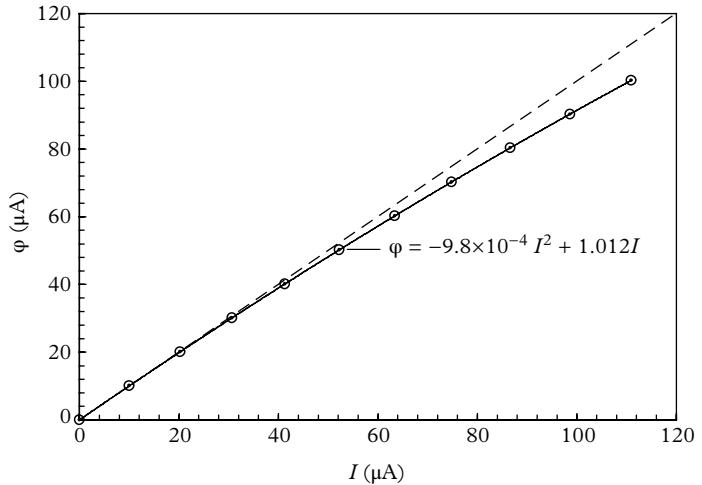


Fig. 9.7. The conversion formula for $I \rightarrow \varphi$ is an ideal way of generating a look-up table for correcting non-linearity. The quadratic equation predicts the actual light input, φ , for any measured anode current, I .

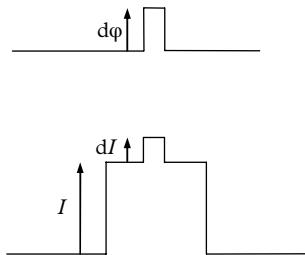


Fig. 9.8. The step-method principle for determining differential non-linearity. This representation is for pulsed operation but the method applies equally to DC signals. When $dI < d\varphi$, the system is said to be non-linear and, when dI is positive, supre-linearity applies.

$K(I)$ is an important parameter because it provides the correction for measured currents using (9.9) and it is plotted along with S in Fig. 9.10. As before, the results of the two methods align well, especially allowing for the relatively wide increment, $d\varphi$, chosen. Integral non-linearity, represented by the function I/φ , and differential linearity, represented by $1/K(I)$, are listed in Table 9.4 and plotted in Fig. 9.11. The single-step method may be adopted, in theory, for pulsed linearity determination, although generating the required waveforms is not trivial. Fenster *et al.* (1973) used a step pulse of 50 μs width in coincidence with one of 30 ms, with measurements taken using an oscilloscope. Such wide pulses would be unsuitable for MCAs, which operate satisfactorily only on pulses of <1 μs.

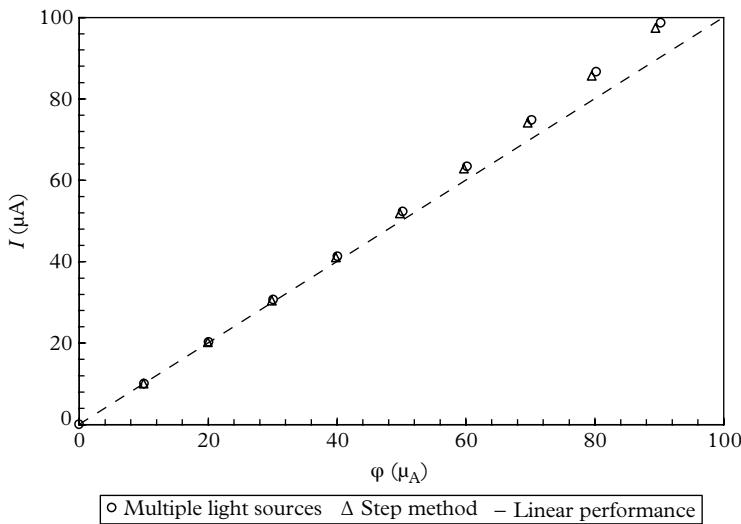


Fig. 9.9. Showing that two detection methods uncover the same degree of non-linearity. Corrected anode currents agree to within 0.1 % with φ . See Tables 9.3 and 9.4.

9

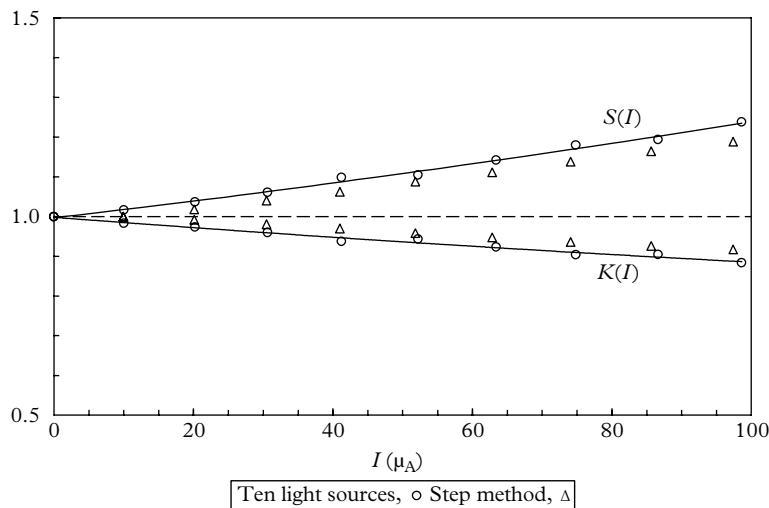


Fig. 9.10. Slope S and correction factor K . The small offset in the two sets of measurements reflects insufficient time to stabilize.

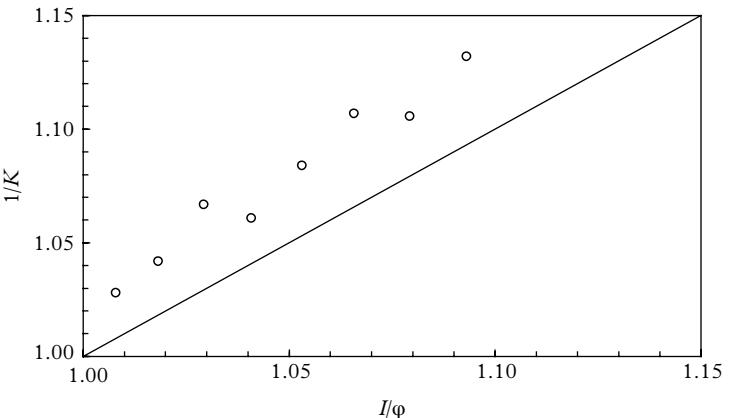


Fig. 9.11. Showing how differential non-linearity exceeds that of integral, highlighting measurement scatter. Data taken from Table 9.3, last two columns.

9.4.2 Ratiometric method

In the ratiometric method, two or more light sources are arranged to give different amplitude signals in a known and fixed ratio to one another. This technique is applicable to DC and pulsed linearity and has been adopted by manufacturers such as Philips (see Piétri 1981, Section 4.3.3.2, and Hamamatsu 2002). A pulse-generator–LED combination is arranged to give a sequence of three pulses, separated in time and in the amplitude ratio $\varphi_1:\varphi_2:\varphi_3 = 6:20:30$ mA, for example. As stipulated in previous considerations, the procedure commences under linear operation, and the LED output signals are therefore set at a suitably low level and low rate, initially. The measurement equipment includes an optical attenuator of fivefold attenuation. When this filter is removed, we have $\varphi'_1:\varphi'_2:\varphi'_3 = 30:100:150$ mA. Non-linearity, ϵ , in the PMT response, shown in Fig. 9.12, is revealed in the peak positions registered by an MCA and it is calculated from

$$\frac{S'_2(\varphi_2)/S'_1(\varphi_1) - S_2(\varphi_2)/S_1(\varphi_1)}{S_2(\varphi_2)/S_1(\varphi_1)} \times 100\%, \quad (9.12)$$

$$\frac{S'_3(\varphi_3)/S'_1(\varphi_1) - S_3(\varphi_3)/S_1(\varphi_1)}{S_3(\varphi_3)/S_1(\varphi_1)} \times 100\%. \quad (9.13)$$

A complete set of readings may be obtained by adding further optical attenuation, or by changing φ ratios. The method described by two manufacturers uses an MCA to determine $S(\varphi)$ and $S'(\varphi)$, although an oscilloscope could be used for this purpose. Peak pulse height (mA) is adopted for the abscissa variable, and channel number (proportional to charge) for the ordinate. Following the

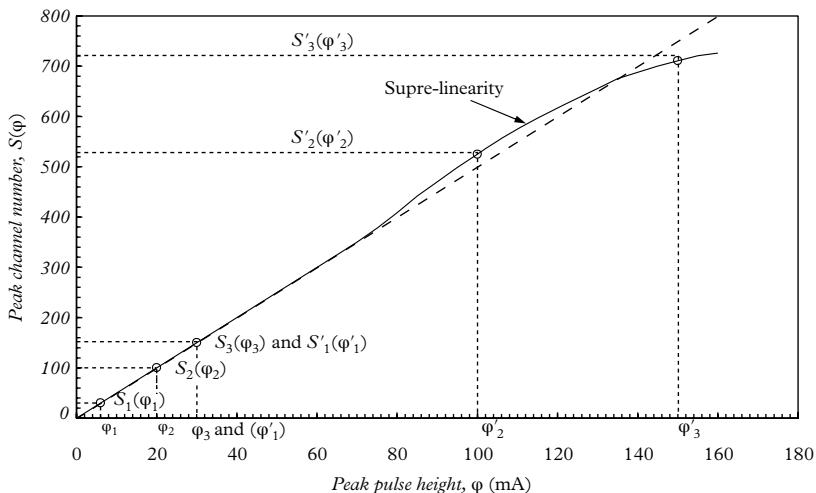


Fig. 9.12. The three-pulse differential linearity test. The solid curve is the actual response, and the dashed curve is the linear response. Other ratio sets for φ may be used, provided φ_3 does not exceed 70 mA—the visual onset of non-linearity.

arguments in 9.3.6 concerning the relationship between pulse height and charge, it would seem preferable to adopt charge for both axes.

9.4.3 Shot noise method

This procedure for measuring differential non-linearity is based on the shot noise formula, which relates the noise on the photocurrent, $\langle i_k^2 \rangle$, to I_k , the mean cathode current. The noise bandwidth, B , enters into the formula as a constant and it need not be known for present purposes. Noise at the anode is enhanced by a constant factor, ENF , and $\langle i_a^2 \rangle$ should therefore still vary linearly with anode current, I_a ; or, equivalently, I_{rms} varies as $I_a^{1/2}$. The attraction of this method lies in its easy realization: it only requires an anode load resistor (110 k Ω for the results in Fig. 9.13); a basic rms DMM or a digital oscilloscope with an rms option; and a random DC light source. Figure 9.13 shows a linear relationship between the variables at low anode currents and a 10% departure at 100 μ A, consistent with errors in the other measurement methods for the same PMT. Noise originating from electrical interference may be a problem but it should enhance low anode current noise only. It is interesting that this is the preferred method for characterizing non-linearity in charge-coupled devices (CCDs) used in the charge integration mode. A plot such as Fig. 9.13 is referred to as a photon-transfer curve, with details of the method to be found in the work of Janesick *et al.* (1987), for example.

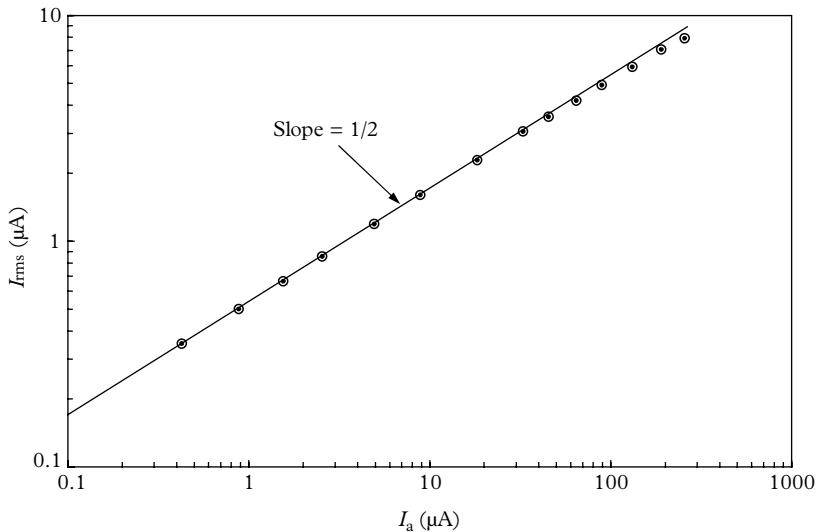


Fig. 9.13. Determination of non-linearity based on the shot noise formula; I_a is the anode current, and I_{rms} is the noise. The straight line has a slope of exactly $1/2$, in accordance with the shot noise formula.

9

9.5 Correcting non-linearity

There are two reasons for measuring non-linearity: (1) as part of the procedure to optimize gain allocation between a PMT and associated electronics, and (2) to accept a degree of non-linear performance suitably corrected. The multiple-light-source method demonstrated in 9.4 is arguably the most precise of all correction methods, particularly by fitting a quadratic to experimental data, with measured (φ, I) coordinates taken from Table 9.3 for substitution into equation (9.14). In the step method, φ has to be deduced following the procedure outlined in equations (9.6) to (9.10), culminating in a correction factor, $K(I)$. Notably, corrected results align closely with those deduced by the multi-source method. By using the equation derived from Fig. 9.6, any given set of measurements of I can be converted to corresponding φ values—a true measure of the light input. In Fig. 9.7 the axes have been reversed, leading to a functional relationship determined by spreadsheet methods:

$$\varphi(I) = -9.8 \times 10^{-4}I^2 + 1.012I. \quad (9.14)$$

9.6 Non-linearity in inorganic crystals

One fundamental requirement of a spectroscopy system is the ability to identify an unknown source of gamma rays by their peak positions. The consequences of a non-linear detector are obvious, and good experimental technique demands investigation of PMT non-linearity and any non-linear scintillator sources. Adaptations of some of the methods outlined in this chapter may be used for this purpose.

There are four sources of non-linearity concerning the use of inorganic crystals

- dependence on the location of the external radiation source with respect to the scintillator
- variability of the light output with respect to the actual location of the gamma capture
- intrinsic non-linearity of the scintillation process: the light output, dL/dE , is a function of E , although the dependence is a weak one
- non-linearity introduced by the PMT.

For further discussion of the first two points, refer to section 3.4 and Figs 3.5 to 3.7.

As will be shown, there is a method for separating the contribution to non-linearity, originating within a PMT, from that due to the intrinsic non-linearity of a scintillator. First, a straightforward investigation is made, given a set of radioactive sources of known energy, and an MCA. The peaks in the distribution correspond to

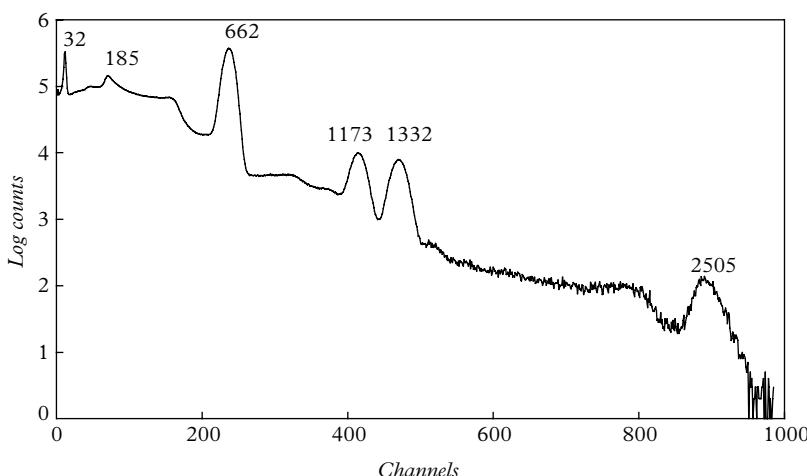


Fig. 9.14. Distribution of gamma rays energies emitted by various radioisotopes. The spectra shows the following photopeaks: the BaK X-ray photopeak, 32 keV; the backscatter peak, ^{137}Cs , 185 keV; the ^{137}Cs photopeak, 662 keV; and the ^{60}Co twin peaks, 1173 keV and 1332 keV, together with a sum peak, 2505 keV. Note a crystal of dimensions $2'' \times 2''$ or larger is desirable in order to detect sufficient 2505 keV gammas.

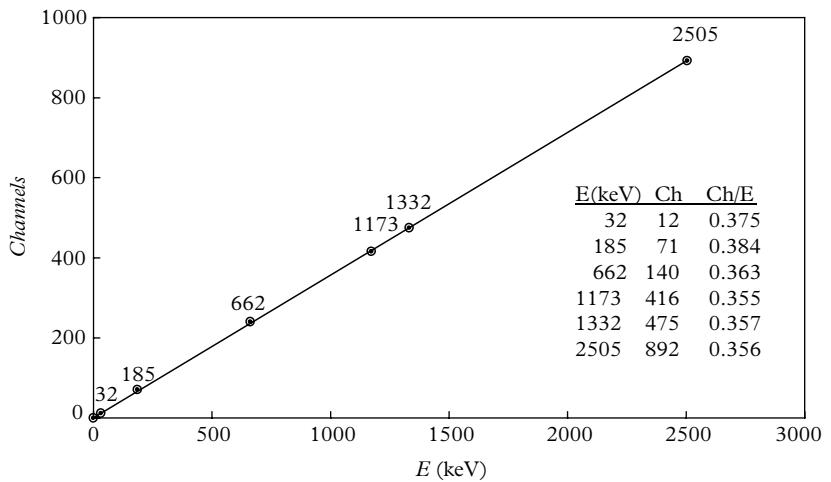


Fig. 9.15. A non-linearity plot of energy, E , versus channel number (proportional to charge) for NaI(Tl). The inset tabular results indicate higher light emission per keV in the region of 100 keV, although there is a lack of precision at these low energies.

9

the total absorption of a gamma ray, and the expectation is a linear relationship between the position of the peaks (in channel numbers) and their associated energy. Energy spectra for a range of standard gamma ray sources are shown in Fig. 9.14, the peak positions of which are plotted in Fig. 9.15. Careful examination of the distribution shows a hint of super-linearity in the readings between 185 and 662 keV; otherwise, linearity applies over the entire energy range displayed. A linear crystal is one for which the peak channel position divided by the corresponding energy is constant. The indication in Fig. 9.15 is that the relative light output is a maximum at about 185 keV, consistent with manufacturers' data.

There is a method for circumventing the intrinsic non-linearity of NaI(Tl), based on the coincident gamma emission of ^{60}Co . The two gammas of energy 1173 and 1332 keV are always emitted in coincidence, but the detection probability is small for a pair. Nevertheless, occasionally both gammas are simultaneously detected, as is evident in Fig. 9.14 by the appearance of a peak at 2505 keV. The twin peaks correspond to the emission of a fixed number of photons (about 50,000) at these energies, and the sum peak equates to exactly the combined output of these two independent light sources. The basis of the method for ascertaining PMT non-linearity stems from (9.1), but crystal non-linearity with energy has no effect in this case. The calibration can be extended to lower channel numbers by varying the degree of optical coupling between the crystal and the PMT. Independence from crystal non-linearity still applies and any remaining deviations refer to the PMT. Other coincident emitters such as ^{22}Na , for example, which emits a gamma of 1275 keV, together with two 511 keV annihilation gammas, may be used in the method under discussion and in the ratiometric one. A useful diagnostic test is to determine the system linearity at a

range of HVs. The sensitivities of the preamplifier, and main amplifier, if applicable, are adjusted at each HV setting to maintain the original overall sensitivity as part of a non-linearity verification for the whole system.

9.7 Summary

The attainment of linear amplification is dependent primarily on the details of the voltage divider and on the magnitude of the HV. It is assumed in all procedures for determining the degree of linear operation that there is a region of performance within which the PMT is linear. Low-level measurements should commence within this region. However, it has been shown that a resistive divider leads to non-linearity under DC operation. This can be minimized but it is always present to some degree. The question of disentangling DC non-linearity in terms of intrinsic and voltage divider effects can be partially answered by using an active divider with the capability of supplying DC anode currents of up to several millamps. It is clear from Fig. 9.16 that intrinsic performance, for this particular PMT, is linear at the 1 % level, for anode currents of up to 100 μ A. This confirms that the source of observed supre-linearity, revealed in Figs 9.6 and 9.7, lies in the voltage divider. It is always beneficial to repeat a set of measurements to gain an indication of the measurement error involved; it is evident from the scatter in Fig. 9.16 that this is of the order of

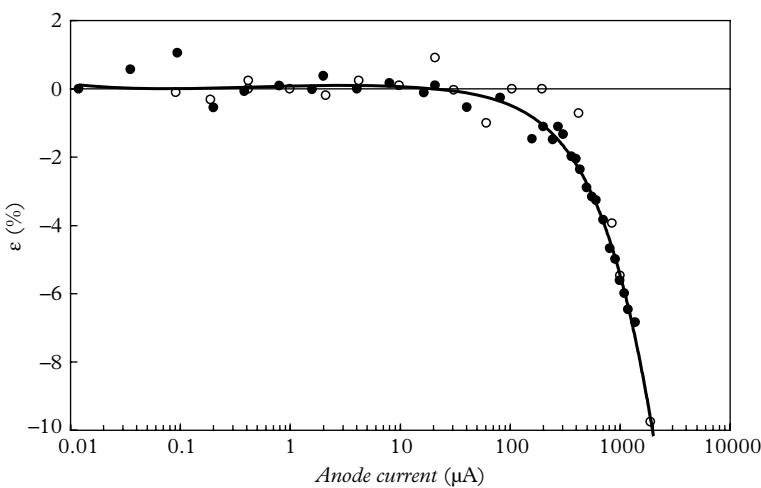


Fig. 9.16. Repeated measurements of ϵ , determined according to (9.1), are consistent with a precision of $\pm 1\%$ but improving on this would require more patience. These results are for a 9107B operated in a fully active divider. The solid line is a best-fit quadratic serving as a visual guide and a correction factor.

$\pm 1\%$. Experimental data are readily corrected for non-linearity by making use of a best-fit quadratic, as is evident in Figs 9.6, 9.7, and 9.16.

The degree of linearity with pulsed sources depends on the type of dynode structure, the voltage divider distribution, the HV, and the mean anode current. If the aim is to quantify intrinsic pulse linearity, then measurements must be made at a repetition rate that creates minimal mean anode current. Pulse height non-linearity can manifest itself as pseudo-improvement in resolution; this is particularly noticeable with dynodes providing limited linearity, such as box-and-grid types. The effect on the spectrum of ^{137}Cs , or isotopes of higher energy, is to squeeze events into a tighter profile, giving false resolution and an incorrect energy allocation. On the other hand, photon counting is based on single-photoelectron pulses suitably amplified but, unlike the case when using spectroscopy pulse height, linearity is relatively unimportant—what matters is that all pulses are counted, and there is even a case for deliberately limiting pulse heights. The presence of a rate effect in photon counting is relatively unimportant, given a PMT with a sufficiently flat plateau characteristic.

The use of separately powered LEDs is important in these investigations because freedom from crosstalk is an essential prerequisite—especially critical in the DC methods. It is essential to follow good earthing practice in constructing test equipment, as eloquently detailed by Ott (1988). Crosstalk between LED signal currents is easily checked by totally absorbing the light from an LED under test (modelling clay is ideal for this purpose) and checking that a neighbouring LED remains unaffected. The light output of this LED, as measured at the anode, must be independent of whether the masked LED is on or off. It is also important to ensure that LEDs are sufficiently energized and free from ‘motor boating’. The light output from an LED should be set several orders of magnitude higher than actually needed. A combination of ND filters, constricting apertures, and the inverse square law can be used to achieve the required light levels. The use of ND filters is acceptable as a means for reducing a light level but their actual attenuation is unlikely to agree with the quoted ND value. The remaining considerations concern possible remedies:

- review allocation of gain between the electronics and the PMT, noting that a small decrease in HV produces a disproportionate reduction in gain
- optimize the voltage divider distribution or consider an active one
- choose a better-suited PMT for the application—perhaps one with fewer stages
- increase the dynamic range by taking a signal off a dynode in addition to one from the anode
- consider the use of an amplifier—a gain of 10 extends the dynamic range by the same multiple
- a silicon detector may be more appropriate at high light levels
- accept the non-linearity, and correct readings
- attenuate the light input—definitely, a last resort

 REFERENCES

- Aramu, F. and Rucci, A. (1967). An improved method of checking the linearity of a photodetector. *Nucl. Instr. and Meth.*, **51**, 153–5.
- De Vincenzi, M. et al. (1984). Experimental study of non-linear effects on photomultiplier gain. *Nucl. Instr. and Meth. in Phys. Res., A*, **226**, 104–12.
- ET Enterprises. (2001). *Understanding Photomultipliers*, Pub. UPMT/01. ET Enterprises Ltd., Uxbridge.
- Fenster, A. et al. (1973). Linearity and fatigue in photomultipliers. *Rev. Sci. Instrum.*, **44**, 689–94.
- Gatti, E. and Piva, F. (1953). A new single channel elementary amplitude discriminator. *Nuovo Cimento*, **10**, 984–5.
- Hamamatsu Photonics K K. (2002). *Photomultiplier Tubes*. Hamamatsu Photonics K K, Hamamatsu City.
- International Electrotechnical Commission. (2010). *BS IEC 60462:2010: Nuclear Instrumentation. Photomultiplier Tubes for Scintillation Counting. Test Procedures*. BSI, London.
- Janesick, J. R., Klassen, K. P., and Elliott, T. (1987). Charge-coupled-device charge-collection efficiency and the photon-transfer technique. *Opt. Eng.*, **26**, 972–80.
- Nicholson, P.W. (1974). *Nuclear Electronics*. John Wiley and Sons, Inc., London.
- Ott, H. W. (1988). *Noise Reduction Techniques in Electronic Systems* (2nd ed.). John Wiley and Sons, Inc., New York.
- Piétri, G. (1981). *Photomultiplicateurs*, Pub. RTC 5482–07–07/1981. R.T.C. La Radiotéhnique-Compelec, Cedex.
- Texas Instruments. (1995). *Understanding Data Converters: Application Report*. Texas Instruments, Dallas, TX. Available at <http://www.ti.com/lit/an/slaa013/slaa013.pdf> (accessed 3 Feb 2017).

10

Collection and counting
efficiency

10.1 Introduction

Collection efficiency refers to the efficacy of a PMT in collecting and propagating photoelectron-initiated signals to the anode. It is an important parameter, particularly for manufacturers, at the design stage of PMTs but also for most users who aim to generate a detectable signal at the anode for every photoelectron released from the photocathode. Absolute measurement of flux is a key topic in light detection but it is difficult to achieve with confidence. Naively, it might appear that making such measurements only requires a PMT with calibrated photosensitivity at the wavelengths of interest. However, obtaining such a device at an affordable cost presents a major difficulty: standards laboratories and manufacturers do not offer PMTs calibrated in the manner of their intended mode of operation. Their preference is to quote QE for illumination based on parallel light confined within a defined area located around the axis of the PMT—hardly a mode of operation to satisfy users. It is clear that hemispherical PMTs pose particular difficulties in this respect. The key parameter, the photocathode QE, is measured with the PMT operated as a photodiode: that is, all electrodes other than the photocathode are connected to the anode. This ensures that photoelectrons emitted by the photocathode, and any secondary electrons released by bombardment of the electrode structure, are collected. There is, in effect, a reduction in QE when a PMT is operated as a photomultiplier, because only a fraction of photoelectrons, F , is collected and propagated from the first to the second dynode and subsequently to the anode. This loss, $1 - F$, is statistically important in low-level-light detection and it is a limiting factor in most attempts at absolute light measurement. In this section, we discuss means for determining collection efficiency, F , for correcting measurements taken at the anode. But the validity of the correction is subject to the uncertainties already mentioned. For manufacturers and users, optimizing F is fundamental to achieving best performance regarding detectivity and resolution.

Light sources offered by standards laboratories can be used to calibrate PMTs in absolute terms, and hence also as an instrument embodying a PMT. A light source is stated to emit a quoted flux of photons at a given wavelength and within a given solid angle. Several of the reported methods used in the determination of F are based on counting all single-photon-initiated events, regardless of their individual size. However, because of the nature of the pulse height spectrum, satisfying the counting requirement is subject to uncertainty. It is customary to operate close to the pedestal of the SER, if an MCA is the choice of instrumentation or, equivalently, close to the start of the plateau where straight photon counting is adopted.

Once calibrated with a light source of known intensity in this manner, the anode signal then represents the absolute light flux incident on the photocathode. On the other hand, if the QE (measured with the PMT operated as a photodiode) is given by the manufacturer, the light flux can only be determined if the collection efficiency is known. Knowledge of these parameters is clearly important in the

planning and modelling stage of a light detector. The requirement might refer to a single PMT in scientific equipment, or a matrix of thousands of PMTs in an astrophysics experiment involving the atmosphere, water, ice, liquid argon, or the scintillator as the light-emitting medium. There are two considerations concerning massive detectors: detectivity and cost, both of which relate to the quantity of PMTs required. Numbers are usually modest in commercial and laboratory instrumentation, and cost may not be the prime consideration in this case.

10.2 Lost photoelectrons

10.2.1 Front end design

Recall from 2.2.1 that QE at a specific wavelength, η_λ , defines the conversion efficiency of photons of wavelength λ to photoelectrons, specified for a PMT operated as a photodiode. Operating a PMT in its amplifying mode presupposes knowledge of the multiplier gain to relate signals measured at the anode to photons detected at the photocathode. The discussion in 5.8 highlights the uncertainties in measuring gain precisely. Adopting unity gain at least leads to consistency among manufacturers. Cathode current I_k , as a function of $\nu(k - d_1)$, is shown in Fig. 10.1 for a light source of fixed intensity. A small leakage current (light switched off), varying linearly with $\nu(k - d_1)$, has been subtracted from the signal

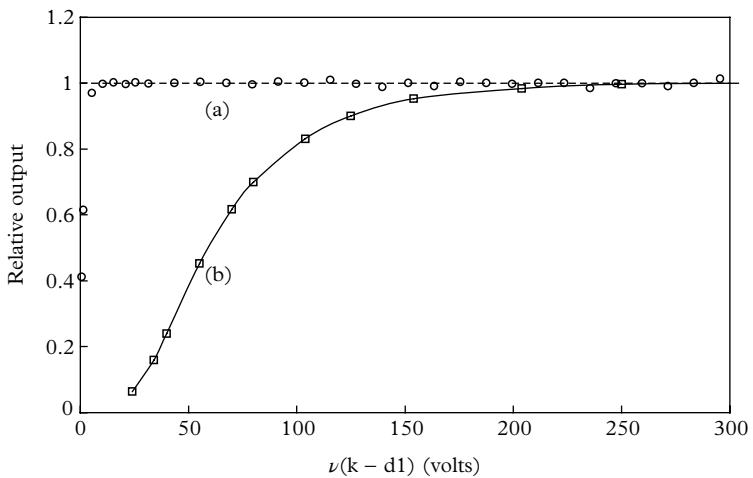


Fig. 10.1. (a) Collection characteristic for a 9214B fast PMT operated as a photodiode. Note that ~ 20 V is sufficient to collect all photoelectrons. This should be compared with PMT operation, (b), where the corresponding figure is ~ 250 V.

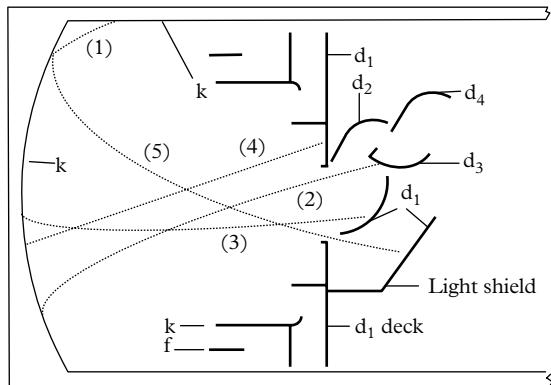


Fig. 10.2. The front end of a fast 50 mm 9214B PMT. The curves labelled (1) to (5), with the exception of (3), represent photoelectron trajectories that fail to propagate. Note that Trajectory (1) does not constitute a loss in collection efficiency since it begins and ends at cathode potential, and no net current flows; it does, however, indicate a loss in QE.

current. The characteristic refers to a PMT with the design shown in Fig. 10.2. It is evident that an applied potential of 20 V is sufficient to collect over 95% of photoelectrons in the photodiode configuration defined in 10.1. In practice, it is inadvisable to operate with $\nu(k - d_1)$ too low, because of enhanced sensitivity to the magnetic field of the earth; poor collection efficiency; reduced resolution; low gain at the first dynode; and sluggish time response. This explains why PMTs are generally operated with $\nu(k - d_1)$ between 100 and 600 V.

10.2.2 Optimal operating conditions

The collection characteristic for operation as a PMT may be determined by using two negative HV supplies: one connected to the cathode, providing $-V_k$ volts, and the other to d_1 , with $-V_{d1}$ held fixed. Single-photoelectron-initiated pulses, generated from a steady light source, appear at the anode once V_k exceeds V_{d1} in magnitude. The size of the output pulses increases with $-V_k$, and the changing rate, determined by photon counting, is indicative of an increase in collection efficiency. The characteristic shown in Fig. 10.1(b) has a slower turn-on compared to that seen with photodiode operation, because of the losses from trajectories such as Trajectories (2), (4), and (5), together with a failure to propagate at low $\nu(k - d_1)$ (see Table 4.3). Note the flattening observed in Fig. 10.1(b) indicates that maximum collection has been achieved at some fixed value for F —always less than unity. The options for detecting single photons are (1) to use an MCA configured to provide differential and integral spectra, and (2) to use an amplifier discriminator, of either fixed or variable threshold. For fixed threshold, counts are recorded as a function of HV. With variable threshold, the HV is

fixed, and the threshold is varied. This is achieved by altering the threshold of the discriminator circuit at fixed HV or by the insertion of a stepped attenuator between the anode and the amplifier discriminator, also at fixed HV. The two modes for determining count rates, fixed HV or variable HV, are not equivalent, as will be demonstrated experimentally in 10.4.2 and theoretically in Appendix A.

10.3 Collection and counting efficiency

10.3.1 Collection efficiency, F

Collection efficiency for single-photon excitation is defined as

$$F = \frac{\text{count rate at the anode}}{\text{rate of photoelectron emission from the photocathode}}. \quad (10.1)$$

The photoelectron rate is deduced from knowledge of the cathode current, $I_k = n_k \eta_\lambda e$. It is inferred because it is too small to measure by direct means, but we shall return to this later. If n_k is the number of photons per second incident on a photocathode, and n_a the corresponding rate of pulses recorded at the anode, with an efficiency of p ,

$$n_a = n_k \times \eta_\lambda \times F \times p. \quad (10.2)$$

Historically, before suitable absolute radiation standards and power meters were readily available, it was usual practice to determine F from (10.2) by extrapolating n_a to zero pulse height, equivalent to taking $p = 1$, making

$$F = \frac{n_a}{n_k \times \eta_\lambda} = \frac{n_a}{I_k/e}. \quad (10.3)$$

10

Gain has a pivotal role in any consideration of absolute measurement and it is important to appreciate the two ways in which it may be specified. PMT gain, G , is the ratio of the anode current, I_a , to the cathode current I_k . The multiplier gain, $\langle g \rangle$, is the ratio of the average charge at the anode, initiated by a photoelectron, to the electronic charge, e ; G is always smaller than $\langle g \rangle$ because of losses embodied in F , whereas $\langle g \rangle$ derives from only those photoelectron contributions that have cascaded and is therefore independent of F . The PMT gain, G , is by definition I_a/I_k but $I_a = n_a \times e \times \langle g \rangle$, from which it follows that

$$G/\langle g \rangle = n_a \times e/I_k = F. \quad (10.4)$$

Equations (10.3) and (10.4) are the basis for determining F , while $\langle g \rangle$ is computed from a pulse height distribution, accumulated using an MCA calibrated in picocoulombs per channel (see 5.7.5); G follows from I_a/I_k , where I_k must be deduced by scaling, using filters. This procedure refers to methods whereby I_k is measured

at a high light level and then scaled by the imposition of calibrated filters to a level that is countable at the anode but too low for current measurement at the cathode.

10.3.2 Anode detection efficiency, ϵ

The ability to assign a figure, ϵ , to a PMT-based detection system has long been a goal of researchers and manufacturers. In the last 20 years, this facility became generally available to experimentalists; for example, astrophysics experiments are based on the use of an absolute photon source, traceable to international standards. As already stated, the parameter ϵ refers to the ratio of the number of pulses detected at the anode to a known number of photons incident on the window of a PMT. Note that extrapolation to zero pulse height is unnecessary, since the user is free to choose the detection threshold in this case. However, the calibration is only valid subject to maintaining the same PMT operating conditions (HV and voltage divider details), and the sensitivity of associated instrumentation (discriminator, ADC, or MCA). Any subsequent changes to the system are likely to invalidate the calibration. This point is worth emphasizing: detection efficiency refers to a specific combination of instrumentation and PMT operating conditions. Furthermore, the area of illumination, the angle of incidence, and the degree of polarization must be duplicated in the calibration procedure.

If a known photon flux, n_k , registers n_a counts/s at the anode then, in units of counts per photon at some set photoelectron equivalent detection level, ϵ is simply

$$\epsilon = n_a / n_k. \quad (10.5)$$

The determination of ϵ stems from knowing the photon flux incident on the photocathode. The QE, η_λ , the counting efficiency p , and the collection efficiency F are implicit in (10.5) and need not be known.

10.4 Measurement techniques

10.4.1 Attenuation using filters

The validity of various methods for determining collection efficiency entered a controversial phase in the late 1960s. Serious differences of opinion existed between two groups of prominent scientists: Foord *et al.* (1969) and Young (1971). The former reported much lower collection efficiencies than other researchers for some common PMT types, particularly the well-established EMI 9558B, S20 PMT. The dynodes in this PMT are of the Venetian blind type, which exhibit a directional response dependent on the approach path of each photoelectron (see 5.6.1). Those moving parallel to a slat bypass d_1 and interact with d_2 , giving rise to an undersized output and, in some instances, no output at all

(see Figs 5.20 and 5.21). The combination of a poor SER and lost photoelectrons is consistent with F in the region of 75%, but not with F of the order of 25%, which was reported by Foord *et al.* (1969). Also, the absence of small pulses reported by Foord *et al.* is not confirmed by others: the P/V ratio for a typical 9558B is 1.2, whereas Foord *et al.* quote a figure of 25—a level even GaP PMTs cannot attain. Claims and counter claims were disputed, and users may wish to draw the obvious conclusions.

The prime difficulty in using (10.3) is one of determining I_k , for it is only a fraction of a picoamp under photon-counting operation. This practical difficulty is overcome, in theory, by measuring cathode current at a high light level, with the PMT operating as a photodiode, and then attenuating the light by a known factor, f , using ND or other types of filter, with the PMT gain restored. The resulting count rate should be such that dead time correction is minimal. The parameter, f , is realized by using the PMT itself to determine the attenuation of a set of filters, with combinations chosen to set f in the region of 10^5 ; I_k is replaced by I_k/f in (10.3). In general, any attenuation procedure, involving multiple reduction steps, enhances uncertainty. However, the majority of authors listed in Table 10.1 have highlighted their particular precautions. For example, under photodiode operation at high light levels, it is important to confirm that I_k is not saturated.

10.4.2 Count rate determination

There is a second source of uncertainty in (10.3) in assigning a number to n_a , since all signal pulses at the anode must be counted. A plateau characteristic, determined using an amplifier discriminator and a variable HV supply, is one way of attempting this. An immediate shortcoming, revealed in Fig. 10.3, is that the plateau region of the curve actually slopes, despite the descriptive terminology. The slope, dn_a/dV , is never zero. This poses a problem in setting the HV, V ; plotting the curve using a logarithmic ordinate does not solve the problem simply by beguiling it. The fact is that plateau characteristics, even for quality PMTs, are better described as regions of minimal slope. Although the distribution in Fig. 10.3 represents such a PMT, it does not provide the precise count rate for substitution in (10.2) to (10.5). The reason for this lies in the choice of transformation variables involved in extrapolation, and this is explored in Appendix A. However, it is feasible to derive an integral count rate curve by changing the sensitivity of the amplifier discriminator while maintaining a fixed HV at, say, 1000 V. A $50\ \Omega$ stepped attenuator is suitable for this purpose, connected between the PMT and the amplifier discriminator. Settings are customarily in decibels but easily converted to a linear scale.

Data taken with an MCA provide a more reliable determination of the value of n_a by extrapolation of the cumulative SER to zero channel number. Noise counts in the lower channels must be replaced by a linear extrapolation based on, say, the first ten noiseless entries, as demonstrated in Fig. 10.4. It is now obvious, by

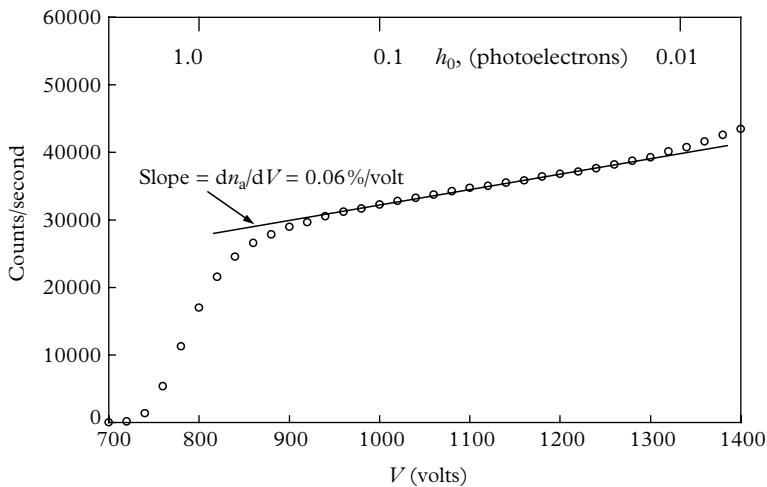


Fig. 10.3. The ‘plateau’ characteristic for a quality PMT always slopes, compromising its use for the determination of n_a . The best that can be claimed is that n_a lies somewhere between 30,000 and 40,000 counts/second. The approximate upper scale on the abscissa reveals the wide dynamic range for the effective threshold, h_0 .

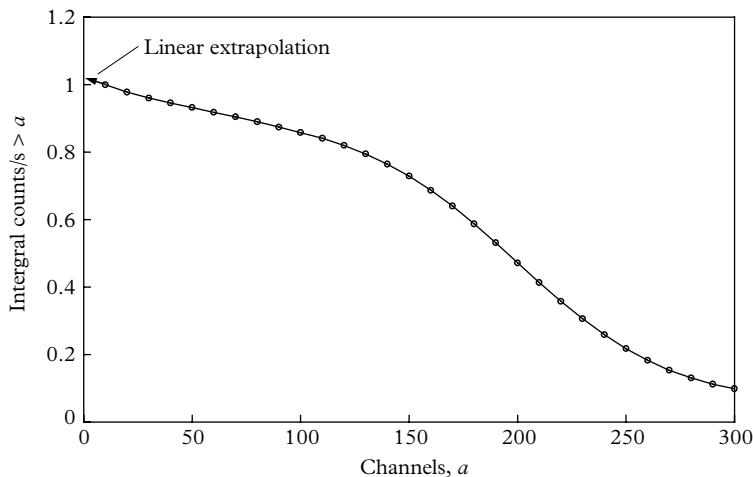


Fig. 10.4. Normalized integral pulse height distribution, $> a$, derived from MCA readings. The uncertainty in determining n_a by the extrapolation shown is judged to be less than 1%.

comparison, that a determination of F based on a plateau characteristic will carry unacceptable uncertainty linked to n_a . The integral count rate, determined from a pulse height distribution at fixed HV, extrapolates in an orderly manner to provide a more reliable estimate of n_a , as is clear in Fig. 10.4.

The PMT gain G , defined in 10.3.1, is given by the ratio of the anode current to the corresponding cathode current, I_a/I_k , and the same light attenuation methods already discussed are applied, to infer I_k . It is preferable to take a series of paired measurements of I_k and I_a using the slope to avoid offsets, which result primarily from leakage currents. These measurements are best performed at an initial gain of $\sim 10^4$ followed by scaling to $\sim 10^7$ by increasing the HV using the relative change in anode current to track the increase in gain. Multiplier gain, φ , is determined at the elevated HV using the relationship $I_a = n_a e \varphi$, or by charge calibration, and F follows from (10.4). The objection to this approach lies in the uncertainties incurred by scaling, as already discussed concerning filters.

10.5 Determination of F

10.5.1 Based on d_1 gain

Certain materials with a high secondary emission coefficient are photosensitive; consequently, the first dynode also exhibits an SER, generated by light transmitted through the photocathode and incident on d_1 . The first stage of multiplication does not contribute to the output in this instance. The mean of the SER initiated at d_1 , as observed at the anode, is consequently reduced by δ_1 , the gain of the first stage. The distribution represented by solid dots in Fig. 10.5 shows two peaks obtained with a narrow collimated beam of light derived from a green LED. A portion of incident light is absorbed by the photocathode, while the remainder illuminates the first dynode. The d_1 SER (circles) alone is recorded by open circuiting the photocathode.

There are two ways in which to specify the gain of the first stage. Measurement of current ratios of I_k and I_{d1} under continuous illumination leads to the parameter, Δ . The procedure for determining Δ requires that the PMT is only partially active by shorting all dynodes beyond d_4 to the anode, as shown in the circuit of Fig. 10.6. This permits reliable measurement of currents I_k and I_{d1} with signals of sufficient magnitude, yet avoiding overloading the remaining gain stages and the anode. The impedance of the multimeter used is $10\text{ M}\Omega$ but knowing its exact resistance is not critical since the same instrument is used to measure the two signal voltages, and the gain depends only on their ratio. Current measurements can therefore be replaced by voltage equivalents V_1 and V_2 . The gain, d_1 , of this mode is determined from $\Delta = (I_k + I_{d1})/I_k$, whereas δ_1 gain requires full PMT performance, in the circuit of Fig. 10.7.

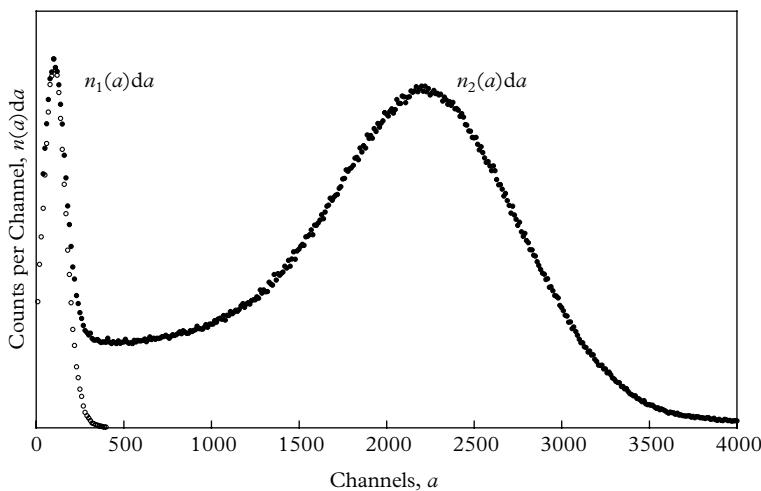


Fig. 10.5. Pulse height distributions for $n_2(a)$ da (●), for an ET Enterprises 9214B operated at $\nu(k - d_1) = 300$ volts, and for $n_1(a)$ da (○), with the latter showing a subpeak. The subpeak disappears on angling the light source, and the SER is flat for channels below 1000 (see also Fig. 5.15 for example). The main peak disappears on open circuiting the photocathode, while the subpeak remains.

Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, 49, 2059–65. Reprinted with permission from OSA ©1970.

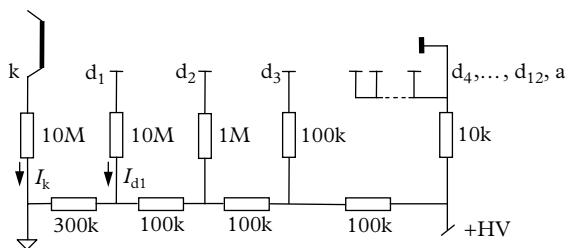


Fig. 10.6. The divider configuration for the measurement of Δ . Monitoring resistors in series with k and d_1 electrodes are decoupled to eliminate noise pick up; $I_k + I_{d1} = I_k\Delta$, from which the Δ may be deduced over a range of $\nu(k - d_1)$ selected by adjustment of HV.

Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, 49, 2059–65. Reprinted with permission from OSA ©1970.

The numerical value for d_1 gain, δ_1 , is based on the SER originating from the photocathode and the SER originating from d_1 . The most probable gain of the first dynode is given by the ratio of the two peak positions, which is 2220/100 in Fig. 10.5. However, for computation of F we need the mean d_1 gain, $\bar{\delta}_1$:

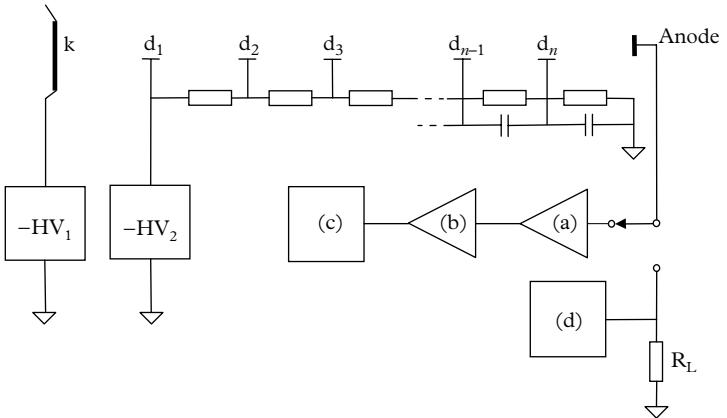


Fig. 10.7. The circuit for determining δ_1 as a function $\nu(k-d_1)$, using two independent power supplies, HV_1 and HV_2 . The circuit elements are as follows: (a) Canberra 2005 charge-sensitive amplifier; (b) Canberra 2022 main amplifier; (c) Canberra MP2-4U MCA; and (d) Wavetek 27XT multimeter, with $R_L = 1 \text{ M}\Omega$ for voltage measurement. Note that either currents or voltages may be measured at (d), since the parameters appear ratiometrically in (10.7).

Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, 49, 2059–65. Reprinted with permission from OSA ©1970.

$$\begin{aligned} \delta_1 &= \bar{a}_2 / \bar{a}_1 \\ &= \frac{\int n_2(a) a da}{\int n_2(a) da} \times \frac{\int n_1(a) da}{\int n_1(a) a da} \quad (10.6) \\ &= \frac{I_2 N_1}{N_2 I_1} \\ &= \frac{V_2 N_1}{N_2 V_1}, \quad (10.7) \end{aligned}$$

10

where the limits for the integrals span the abscissa (Channels 10 to 500).

Distributions $n_1(a)da$ and $n_2(a)da$ in Fig. 10.8 are pulse height spectra for the Δ SER and the δ_1 SER, respectively. We note that $\int n_2(a)ada$ and $\int n_1(a)ada$ are proportional to I_2 and I_1 , leading to (10.7). The expression $\int n_2(a)da = N_2$, integrated from Channel 10 to the end of the spectrum, is the cumulative count rate recorded with an MCA. This distribution is normalized after dividing by N_2 and is shown in Fig. 10.9, together with $\int n_1(a)da = N_1$, which is also normalized.

Collection efficiency is given by the ratio of the gain parameters measured over a range of $\nu(k - d_1)$ voltages:

$$F = \Delta / \delta_1, \quad (10.8)$$

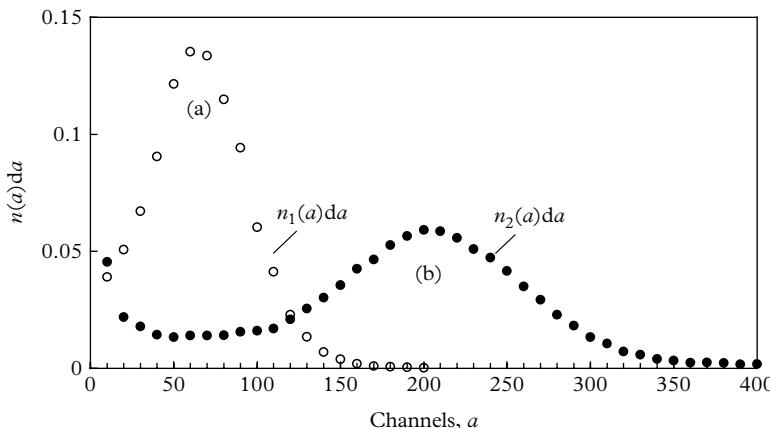


Fig. 10.8. (a) The d_1 -SER and (b) the SER for a R6095P, 28 mm PMT measured with $\nu(k - d_1) = 100$ V. Note the abscissas for the SER have been scaled down by a factor of 3 for presentation purposes (the peak of the SER is actually at Channel 600).

Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, 49, 2059–65. Reprinted with permission from OSA ©1970.

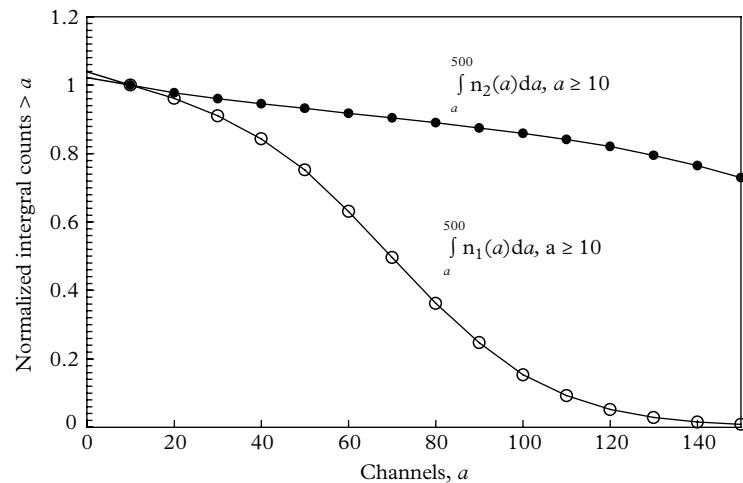


Fig. 10.9. Normalized integral pulse height distributions (normalized at Channel 10) derived from Fig. 10.8. The uncertainty in determining N_1 and N_2 by extrapolation is of the order of 1 %.

Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, 49, 2059–65. Reprinted with permission from OSA ©1970.

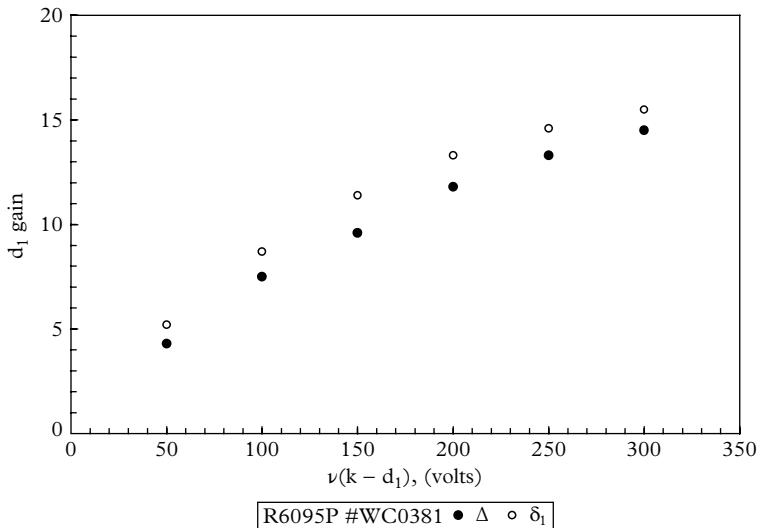


Fig. 10.10. The parameter Δ , measured under DC conditions with green light (●) and with δ_1 determined from (10.7) (○); F is the ratio of the gain pairs according to (10.8). Note there is only a small dependence on the wavelength of the light.

Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, 49, 2059–65. Reprinted with permission from OSA ©1970.

with the Δ and δ_1 values taken from the readings in Fig. 10.10. Error bars shown in Fig. 10.11 are systematic only. The R6094 and R6095 30 mm diameter PMTs are seen to collect photoelectrons with high efficiency; this is attributed to their relatively large d_1 aperture of $\sim 1 \text{ cm}^2$. On the other hand, the 9214B has a considerably smaller aperture than the R6095, in addition to a photocathode area three times bigger.

The methods covered so far are aimed at correcting manufacturer's QE data (always a reduction) to allow for the loss of photoelectrons, principally between the photocathode and the first dynode. Knowing F and given the manufacturer's standard the QE calibration (photodiode operation) leads to a figure for the incident light flux on the photocathode, albeit with considerable uncertainty. No allowance for reflections from surfaces and interfaces related to the photocathode need be made, since these losses are already included in QE calibration. It has been demonstrated that any experimental method that relies on plateau counts can only assign a collection efficiency of considerable uncertainty.

10.5.2 Shot noise

The shot noise formulae for noise, N , and the signal-to-noise ratio S/N , stated in the usual nomenclature, are, respectively,

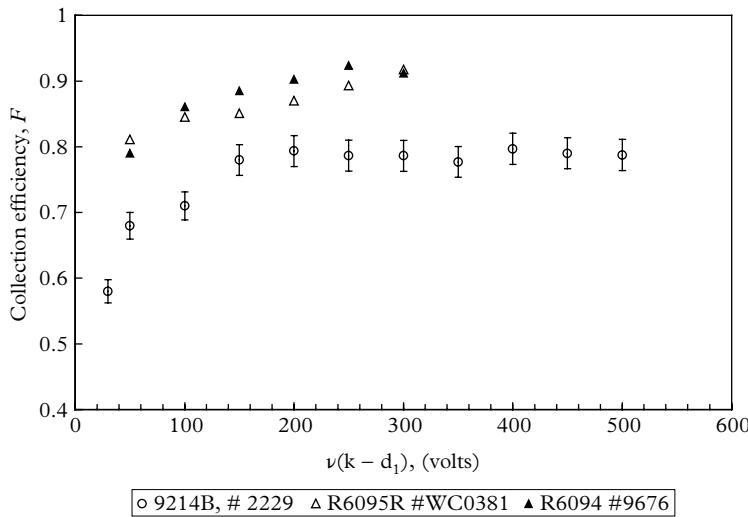


Fig. 10.11 Collection efficiency as a function of $\nu(k - d_1)$ for a fast PMT (9214B) and two 30 mm R6094 and R6095 types. Error bars apply to the other sets of data.

Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, 49, 2059–65. Reprinted with permission from OSA ©1970.

10

$$N = (2eBI_k)^{1/2}$$

and

$$S/N = I_k / (2eBI_k)^{1/2}.$$

Two refinements must be made to these formulations to predict N and the S/N at the anode: allowance for photoelectrons that fail to propagate; and a contribution to noise from the multiplier. Next, I_k is replaced by FI_k , and the inclusion of a multiplier noise factor, $1 + \text{var}(g)/\langle g \rangle^2$ predicts the total noise at the anode:

$$N^2 = \langle I_a^2 \rangle = 2eBFI_k \langle g \rangle^2.$$

Noting that noise acts on the surviving cathode current and not on I_k , and recalling that

$$\langle g^2 \rangle = \langle g \rangle^2 \left(1 + \text{var}(g) / \langle g \rangle^2 \right),$$

therefore

$$\langle I_a^2 \rangle = 2eBF \langle I_k \rangle \langle g \rangle^2 \left(1 + \text{var}(g) / \langle g \rangle^2 \right)$$

and

$$F = \langle I_a^2 \rangle / \{ 2eBI_k \langle g \rangle^2 [1 + \text{var}(g) / \langle g \rangle^2] \}. \quad (10.9)$$

The signal-to-noise ratio is then given by

$$(S/N)^2 = \frac{(FI_k\langle g \rangle)^2}{2eBFI_k\langle g \rangle^2(1 + \text{var}(g)/\langle g \rangle^2)}$$

$$= \frac{FI_k}{2eB(1 + \text{var}(g)/\langle g \rangle^2)},$$

from which F follows as

$$F = (S/N)^2 \frac{2eB(1 + \text{var}(g)/\langle g \rangle^2)}{I_k}. \quad (10.10)$$

Young (1971) makes two important points: the expression in (10.9) contains $\langle g \rangle^2$, making F sensitive to any error in $\langle g \rangle$; note that $I_k = I_a/Fg$ but its substitution in (10.10) would lead to a circular argument. To make use of (10.9), $\langle I_a^2 \rangle$ is measured with a true rms meter of known bandwidth I_k is deduced using calibrated filters, and $\langle g \rangle$ must be predetermined by, for example, using one of the methods outlined in 5.8. Similarly, to make use of (10.10), the signal-to-noise ratio is measured with a true rms meter, and I_k is deduced. As is now widely recognized, it is preferable to determine I_k by using calibrated filters, rather than attempting to measure it simultaneously with anode current. In any case, both of these procedures are of questionable validity.

10.5.3 The bi-photonic field

A bi-photon source derived from spontaneous parametric scattering (SPS) offers a means for calibrating the absolute efficiency of light detectors by providing pairs of coincident photons. One of the earliest reports is due to Burnham and Weinberg (1970) subsequently followed by Malygin *et al.* (1981, 1985), Sergienko and Penin (1986), Penin and Sergienko (1991), and Migdall *et al.* (1995/6). The excitation wavelength adopted by Burnham and Weinberg (1970), for example, is 325 nm, which is converted to a pair of coincident photons of wavelengths 633 nm and 688 nm. QE calibration is generally feasible at wavelengths between 600 and 800 nm and hence suitable for S20 and III-V PMTs.

The method is straightforward and is described as follows. Two detectors are exposed to coincident photon fluxes of N_1 at a wavelength λ_1 , and N_2 at λ_2 , respectively. If the detector QEs are η_1 and η_2 the rate of output pulses from each detector is

$$R_1 = \eta_1 N_1$$

and

$$R_2 = \eta_2 N_2$$

and the rate of coincident pulses, M , is

$$M = \eta_1 R_2$$

since the first detector collects the partner of any photon detected by the second detector. Similarly, $M = \eta_2 R_1$. The respective QEs are

$$\eta_1 = M/R_2$$

and

$$\eta_2 = M/R_1.$$

The opening angle of the photon pairs is small, requiring a long throw to achieve sufficient separation. This makes the equipment unwieldy, which, together with the cost of the light sources, may explain why these systems are not more ubiquitous.

10.5.4 Calibrated lamp

An example of the use of a calibrated deuterium lamp, of certified spectral radiance (Institut National de Metrologie, CNAM, Paris) is described by Besson *et al.* (1994) for investigations with UV light. This type of lamp is extremely bright by PMT standards and requires considerable attenuation with filters and apertures to attain photon-counting levels. The rate of photon collection, N , is known from the calibration, and F is given by:

$$F = \frac{I_a}{\eta N e g}$$

Making measurements using lamps is undoubtedly the most difficult of all the methods available for determining F . The difficulty is exacerbated when measurements are made in the UV and VUV ranges. Obvious drawbacks are the sizeable light attenuation factor necessary, and having to work with invisible UV light.

10.5.5 Cerenkov light source

As far as the author is aware, the use of such a source has been confined to the calibration of massive liquid and underwater experiments. Although there are difficulties in creating a ‘standard candle’, the advantages of this type of source are that it requires no power supply, exhibits long-term stability, is affordable, and is portable. The method adopted by Boardman *et al.* (1994) has the merit of conforming closely to the intended application. The SNO experiments detect Cerenkov radiation from relativistic electrons in heavy water. A 50 mm diameter water-filled glass bulb, doped with a $^{90}\text{Sr}/^{90}\text{Y}$ beta emitter, ultimately provides a stable and portable light source with an acceptable single-photon emission rate. The efficiency for collecting single photons at a distant photocathode needs to be very low but achieving this is difficult because Cerenkov light is directional and multi-photon. Furthermore, the light intensity has a $1/\lambda^2$ spectrum which may complicate the method in single-wavelength applications.

A desirable characteristic of this type of source is that the light flux may be calculated from theory.

Accepting the computation for the Cerenkov yield and all the corrections, the detection rate at the anode is

$$R = 2\pi c A \Omega \int_{300}^{700} \eta(\lambda) T(\lambda) \frac{dN/d\omega}{\lambda^2} d\lambda, \quad (10.11)$$

where:

ϵ is the probability that a single photoelectron will register as a count

c is the speed of light

A is the source activity

Ω is the solid angle subtended by the photocathode at the source

$\eta(\lambda)$ is the QE of the photocathode

$T(\lambda)$ is the transmission of a short wavelength cut-off filter

$dN/d\omega$ is the mean source intensity per beta decay

The difference between the measured rate of events and the parameter R leads to the detection efficiency. This gives the effective QE, which includes the collection efficiency, F , and counting losses. Unlike previously mentioned methods, count rates are not extrapolated to zero pulse height. In a follow-on study by the same group, Biller (1999) adopted a threshold of 0.25 pe. This had the effect of reducing the peak QE of a typical PMT from about 28% to a counting efficiency of $\sim 12\%$. Roughly half the loss in sensitivity is due to F , and the remainder to threshold losses. Their method is to calibrate what comes out of the anode and not what is lost between cathode and the first dynode. The advantage of using a calibrated light source should now be obvious: it provides the number of photons incident on the photocathode, whether pulsed or continuous, with the proviso that the sensitivity of the detection electronics must not be changed subsequently.

10

10.5.6 Use of a power meter

The method described by Lakes and Poultney (1971) is based on the use of a calibrated power meter (Spectra Physics 401), suitably attenuated to give $\sim 10^5$ counts/s, N , at the anode. The same power meter is used to measure the attenuation of a set of filters required for reducing the intensity of the laser beam to single-photon levels. The degree of attenuation is $\sim 10^9$ but must be accurately known to infer n . Collection efficiency follows by measuring N for substitution in $F = N/m\eta$. The method suffers from the same propagation errors that trouble other methods based on attenuation.

10.5.7 Calibrated silicon photodiodes

Following on from the work of Boardman *et al.* (1994), to produce a light source of known photon rate, a group which included the National Physical Laboratory (NPL) was formed (Biller *et al.* 1999). The key feature of this investigation was the use of 1 cm² photodiodes, type Hamamatsu 1227 and 1237. The method involved focussing a monochromatic beam into a 50 mm diameter, two-exit-port, integrating sphere. One port provided illumination of a distant PMT under test, while the other monitored the input light intensity. The exceptional linearity offered by the monitor diode is used to lower the intensity of the light input to the sphere by a known factor. This is achieved by interposing ND filters before the integrating sphere to achieve an attenuation of $\sim 10^5$ but known precisely. The key feature of this procedure is that the end result does not depend on the quoted ND number but is derived from the change in the monitor diode reading. The absolute flux of light incident on the photocathode was determined by the reference photodiode located between the sphere and the PMT. This diode was calibrated against the NPL standard cryogenic radiometer, as described in 2.3.4.

A feature of all calibration methods, for which the input light flux is known, is that it removes the requirement to extrapolate integral anode counts to the zero discriminator level. In the method of Biller *et al.* (1999), for example, a counting threshold of 0.25 photoelectron equivalent was chosen to locate the operating point in the approximate position of the valley of the SER. The integral count rate above this threshold is thus identified with a known photon flux. The choice of such a discrimination level ensures that 70 to 80% of the counts are included, thus limiting loss of statistical information. Furthermore, the valley of an SER is the most stable operating point in the distribution and is consequently a preferred operating point. The detection efficiency is a function of wavelength and combines QE, collection efficiency, F , and the threshold efficiency—that is, the probability that a pulse is detectable at the anode. These authors present efficiency curves that have the same shape as the manufacturers' QE curves but are reduced in area by about 50%.

The results presented by Birenbaum and Scarl (1973) for a C31000F, GaP PMT are worth examining in the light of the extrapolation procedure just outlined. They determined F for a PMT with a rather poor SER showing a high level of small pulses. Consequently, the plateau characteristic of counts against HV has a slope in excess of 10% per 100 V, and the zero threshold count rate cannot be revealed. Making use of a calibrated RCA 6342, they were able to quote QE and counting efficiency.

The method adopted by the High Resolution Fly's Eye Experiment (HiRes; Bird *et al.* 1994) is novel since the researchers do not count single photons, although their ultimate calibration is in picocoulombs per incident photon. PMTs are pointed at the night sky in this experiment, and the background from starlight restricts the gain to 2×10^5 , which is about two orders of magnitude lower than that recommended for photon counting. Bird *et al.* make use of a combination of S1126-8BQ silicon

photodiodes, in the 100% trapped configuration (see 2.3.5), and low-noise transimpedance ($1\text{ G}\Omega$) amplifiers; the combined silicon-detector–amplifier combination was calibrated by NIST for absolute sensitivity. The calibration is based on measuring cathode current and anode current successively without recourse to filters. Although previously described as undesirable by the present author at photon-counting levels, this technique is acceptable at low gain.

10.6 Summary and conclusions

The entries in Table 10.1 are representative of measurements made specifically for F . This parameter is undoubtedly of interest to designers of PMTs, whereas users would prefer to know the detection efficiency: that is, the rate of events detected at the output of some chosen instrumentation, divided by the input flux. This is achieved by Biller *et al.* (1999) in their Figure 4, where detection efficiencies are quoted at a judicious discriminator threshold of one-quarter of a photoelectron. It follows from their QE efficiency distributions that this corresponds to $F_p \approx 50\%$, but this figure does not enter into the absolute calibration. Similarly, Bird *et al.* (1994), through the use of

Table 10.1 Reported values for F .

Group	Method	PMT type	F
Foord <i>et al.</i> (1969)	Noise in signal	EMI 9558	0.25
	Current measurement	EMI 9658	0.29
	Current measurement	ITT FW 130	0.18
Lakes and Poultnay (1971)	Optical attenuation	RCA C31000E	0.75
	Optical attenuation	EMI 9558	0.80
	Optical attenuation	Amperex 56TVP	0.71
Young and Schild (1971)	Optical attenuation	EMI 9558	0.75
Birenbaum and Scarl (1973)	Optical attenuation	RCA C31000F	0.61
Coates (1973)	Optical attenuation	EMI 9558	0.58
	Optical attenuation	RCA 8852	0.98
	Optical attenuation	RCA 8850	0.95
Besson <i>et al.</i> (1994)	Deuterium lamp	Photonis XP2020	0.80
Wright (2010)	SERs	EMI 9214	0.78
	SERs	HPK R6095	0.90

calibrated silicon diodes, have succeeded in calibrating the HiRes experiment in terms of picocoulombs per photon, specifically at low PMT gain.

Reliable measurements of collection efficiency are difficult to obtain, for the following reasons:

- Methods based on knowing the photocurrent are subject to severe difficulties in its direct measurement. It is immeasurable under photon-counting conditions and must be deduced from filter attenuation factors measured in situ. This method always presents difficulties.
- Where the anode count rate enters into the determination, it is necessary to establish this by extrapolation—a subjective procedure.
- A correction factor $(1 + \gamma n)$ has been suggested by Coates (1973) to allow for afterpulses, where γ is the afterpulse rate/signal rate, and n the average afterpulse size in photoelectrons equivalent. Establishing the magnitude of this correction is uncertain.
- Drawing of anode current must not compromise the gain stability provided by the voltage divider.
- Exact duplication of application conditions, such as angular dependence of incident radiation, polarization, the voltage divider, dead time corrections, and presence of magnetic fields, is important.
- To date, all determinations are based on asynchronous detection, whereas a synchronous mode could be configured to exclude afterpulses, from count rates but not from current measurement. Note the shape of the SER depends on the mode adopted.
- It is easily verified that F and $\langle g \rangle$ are independent of signal level, provided the PMT is operating linearly.

REFERENCES

- Besson, P., et al. (1994). Measurement of photomultiplier quantum efficiency. *Nucl. Instr. and Meth. in Phys. Res. A*, **334**, 435–7.
- Biller, S. D. et al. (1999). Measurement of photomultiplier single photon counting efficiency for the Sudbury Neutrino Observatory. *Nucl. Instr. and Meth. in Phys. Res. A*, **432**, 364–73.
- Bird, D., et al. (1994). The calibration of the absolute sensitivity of photomultiplier tubes in the High Resolution Fly's Eye Detector. *Nucl. Instr. and Meth. in Phys. Res. A*, **349**, 592–9.
- Birenbaum, L. and Scarl, D. B. (1973). Photomultiplier single photon counting efficiency. *Appl. Opt.*, **12**, 519–21.
- Boardman, R. J., et al. (1994). A Cerenkov radiation source for photomultiplier calibration. *Nucl. Instr. and Meth. in Phys. Res. A*, **345**, 356–66.

- Burnham, D. C. and Weinberg, D. L. (1970). Observation of simultaneity in parametric production of optical photon pairs. *Phys. Rev. Lett.*, **25**, 84–7.
- Coates, P. B. (1973). Photomultiplier collection efficiencies and nonpoissonian pulse height distributions. *J. Phys. D: Appl. Phys.*, **6**, 153–63.
- Foord, R., et al. (1969). Use of photomultiplier tubes for photon counting *Appl. Opt.*, **8**, 1975–89.
- Lakes, R. S. and Poultney, S. K. (1971). Direct measurement of the quantum counting efficiency of RCA C31000E/F photomultipliers at 6328 Å. *Appl. Opt.*, **10**, 797–800.
- Malygin, A. A., Penin, A. N., and Sergienko, A. V. (1981). Absolute calibration of the sensitivity of photodetectors using a biphotonic field. *JETP Lett.*, **33**, 477–80.
- Malygin, A. A., Penin, A. N., and Sergienko, A. V. (1985). Spatiotemporal grouping of photons in spontaneous parametric scattering of light. *Sov. Phys. Dokl.*, **30**, 227–9.
- Migdall, A. L. et al. (1995/6). Absolute detector quantum-efficiency measurements using correlated photons. *Metrologica*, **32**, 479–83.
- Penin, A. N. and Sergienko, A. V. (1991). Absolute standardless calibration of photodetectors based on quantum two-photon fields. *Appl. Opt.*, **30**, 3582–8.
- Sergienko, A. V. and Penin, A. N. (1986). Absolute calibration of analog photodetectors with a biphoton field. *Sov. Tech. Phys. Lett.*, **12**, 328–9.
- Wright, A. G. (2010). Method for the determination of photomultiplier collection efficiency, *F. Appl. Opt.*, **49**, 2059–65.
- Young, A. T. (1971). Use of photomultiplier tubes for photon counting. *Appl. Opt.*, **10**, 1681–3.
- Young, A. T. and Schild, R. E. (1971). Photoelectron collection efficiency in photomultipliers. *Appl. Opt.*, **10**, 1668–72.

11

Signal-induced background

11.1 Introduction

Signal-induced background has a time dependence that distinguishes it from the sources discussed in Chapter 6. When background events have the characteristics of signal, they may be described as unwanted signal; we also have a contribution from background-induced background, showing time dependence. Signal-induced events refer to a progression in which a signal generates a subsequent one, correlated in time to the initial detection. The timescale for correlated background ranges from nanoseconds to days and it is appropriate to categorize such phenomena in terms of their time of occurrence, relative to an initiating signal.

- (1) The earliest occurring timing artefact is a prepulse, which is caused by a photon that penetrates the cathode, producing a photoelectron at d_1 and an output signal that occurs earlier than the cathode-to-anode transit time. The transit time of prepulses to the anode is reduced by the cathode-to- d_1 propagation time.
- (2) Paired-electron-initiated events occur with a time separation ranging from a few nanoseconds to about 100 ns. A fraction of late pulses relate to the cathode-to- d_1 transit time, and others to the propagation time between cathode and anode.
- (3) The next category ranges over time intervals of 100 ns to approximately 10 μ s: these timescales are consistent with ion transit times between d_1 and the cathode. They are referred to as afterpulses.
- (4) There is a long-lived decay signature consisting of single-photoelectron events extending to days, caused by exposure to bright light or high levels of radiation.

The artefacts mentioned above affect all applications where measuring time differences is of primary concern. In the section that follows, we investigate the influence that these timing aberrations have on established instrumentation and methods.

11.2 Timing instrumentation

11.2.1 Amplifier discriminators

An amplifier discriminator of fixed threshold, shown schematically in Fig. 11.1, is a prerequisite for making timing measurements of the four listed types. Its role is to produce a standard digital output pulse for every input pulse exceeding a set threshold. TTL and ECL were the preferred protocols for the digital output in

the sixties but current units tend to use the low-voltage-differential (LVD) logic levels for the signalling output format. Amplifier discriminators are available in three forms: board mounted, compact boxed units, and NIM units. The latter comprise a fast NIM amplifier coupled to an NIM discriminator of 50 mV sensitivity. It has been shown, for photon counting (7.5.7), that a threshold of 2 mV is suitable if used with a fast PMT operated at a multiplier gain of the order of 10^7 .

11.2.2 Configurations for afterpulse measurements

Afterpulse measurements can be made by combining a pulse generator with the amplifier discriminator shown in Fig. 11.1 to realize the waveforms depicted in Fig. 11.2. The output signal from the amplifier discriminator initiates a 10 μs delayed pulse from the pulse generator. Any afterpulse falling within the 10 μs gate generates Waveform (c), and is counted. The ratio of the rates from Waveforms (c) and (a) gives the percentage of counts that produce correlated signals within this time range. If random photons are used for excitation, the rate must be

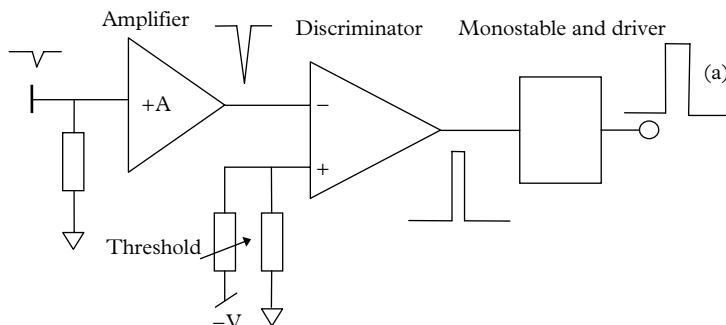


Fig. 11.1. An amplifier discriminator provides the basic means for generating a plateau characteristic. The addition of logic circuitry produces a unit for making timing measurements.

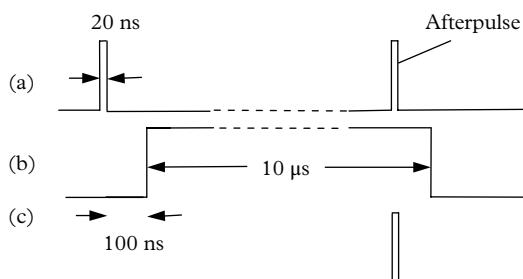


Fig. 11.2. A timing diagram for the measurement of the total afterpulse rate. The delay of 100 ns ensures that Type 2 contributions are excluded.

sufficiently low to minimize chance coincidences within the 10 μs gate. A synchronous single-photon source of sufficient purity is essential. The test method is therefore necessarily inefficient to ensure that only a small percentage of laser flashes result in the generation of a single photoelectron. Consequently, the probability of producing more than a single-photoelectron-initiating signal is negligible (see 4.1.1). A synchronous single-photon source is considerably more desirable, since all recorded delayed events are signal related. Nowadays, a digital oscilloscope is easily programmed to perform the functions indicated in Fig. 11.2. In practice, Waveform (c) will sometimes also produce an afterpulse but the occurrence of these double events is rare. An indication of time structure can be gained by reducing the width of Waveform (b) to, say, 50 ns, followed by scanning the delay gate, but this is laborious. An autocorrelator and an amplifier discriminator acting on random single photons provide an alternative method for recording afterpulse time distributions, particularly by embracing DSP.

11.2.3 Stop–start instrumentation

The limitation of the scheme in 11.2.2 for determining the fine structure of correlated events should be evident. Furthermore, the arrangement cannot record prepulses. The stop–start instrumentation shown in Fig. 11.3 is suitable for study of Categories (1), (2), and (3) in 11.1, with output in the form of a differential time distribution of fine detail. The set-up depicted in Fig. 11.3 is typical of stop–start arrangements, with most investigators preferring synchronous operation with a pulsed laser or fast LED as the source. As mentioned in 11.2.2, a single-photon source of sufficient purity, whether synchronous or otherwise, is essential. The test method is therefore necessarily inefficient but by interchanging the leads of the start and stop inputs, together with resetting the delay, removes a plethora of null events. The TAC converts a digital pulse of variable length to one of

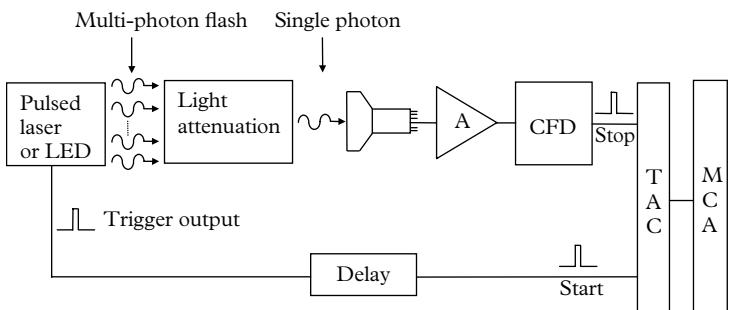


Fig. 11.3. Transit-time distribution apparatus based on delayed coincidence method of Bollinger and Thomas (1961). The original equipment derived the trigger from a scintillation counter; note that a TAC is also called a time-to-digital converter (TDC).

corresponding height and fixed width. An MCA allocates every input to the appropriate size bin, accumulating a spectrum in so doing. It is customary to use a fast amplifier, of gain greater than 10, in combination with a CFD of 50 mV threshold. PMT gain is adjusted to achieve a threshold in the region of 0.1 and 0.25 photoelectrons equivalent—ideally, the same threshold as in the intended application. Operating details under which transit-time distributions are measured are important and, as for other characterizations of PMT behaviour, they should be compatible with the application. Furthermore, timing distributions change depending on whether operation is synchronous or asynchronous. Prepulse time distributions can be recorded with the synchronous arrangement of Fig. 11.3.

11.3 Prepulses and late pulses

An ideal PMT is one that produces a single pulse for each photoelectron released. Signals at the anode are characterized by a mean transit time, t_t , relative to the time of creation of the photoelectron, but subject to statistical spread. This delay is influenced by several factors, among which the diameter of the PMT is a major consideration (path lengths are necessarily long). For a 200 mm diameter hemispherical PMT, such as the 9351B, t_t is ~ 75 ns, whereas it is only ~ 5 ns for a metal-can PMT of diameter 16 mm (R7400 series). Other contributions to transit time depend on the type and number of dynodes, voltage divider design, and operating voltage. Manufacturers quote the mean transit time, and the dispersion in this parameter, for all products. It is common practice to fit a Gaussian distribution to the main peak, shown by the dashed wings in Fig. 11.4; all events lying outside these confines are taken as artefacts. Furthermore, there are the occasional events for which t_t lies significantly outside the expected statistical boundaries by those signals arriving either earlier or later than the mean transit time (the prepulses and late pulses). Consequently, applications based on timing, such as massive neutrino detectors, stop–start systems, LIDAR, fluorescent lifetime, and autocorrelation studies may show features that are artefacts of undesirable PMT behaviour. Awareness of these phenomena goes back 50 years, as reported by Bonitz *et al.* (1964) on prepulses, and Stevens and Longworth (1972), concerning late pulses. Understanding the origins and means for suppressing these events is of critical importance in massive astrophysics and other experiments that require timing information to achieve detailed event reconstruction. Lubsandorzhev *et al.* (2006) have measured transit-time parameters for 14 different PMT types, while Baldini *et al.* (1996) (CHOOZ) and Ianni *et al.* (2005) (Borexino) have results for 250 and 2200 hemispherical PMTs, respectively. These PMTs are of the same generic type, and the results show consistency in performance. All of these research groups provide convincing explanations regarding the physics of the processes involved. Figure 11.4 is a typical transit-time

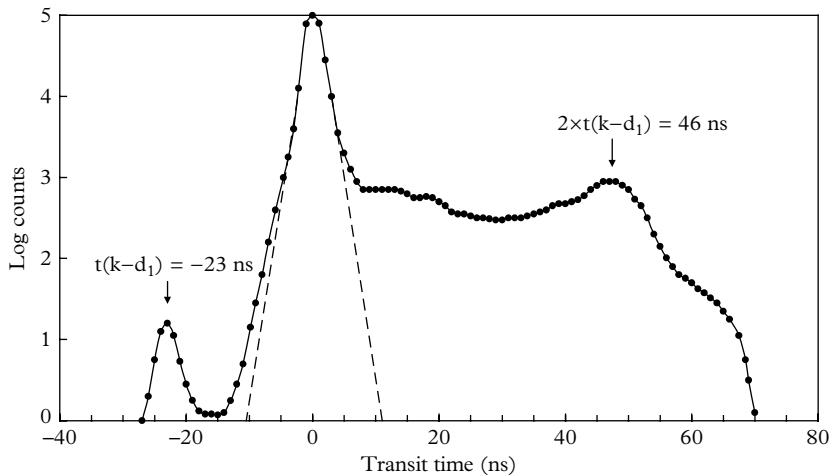


Fig. 11.4. Single-photoelectron transit-time distribution for an 8" hemispherical ETL 9351 B PMT. The distribution was measured with a stop–start system similar to that in Fig. 11.3. The choice of a log scale for the ordinate reveals details that would otherwise be missed.

dispersion response for single-photon excitation of a hemispherical PMT; the mean transit time is 70 ns, corresponding to the main peak in the timing distribution, taken as time zero. The notable features are the prepulse events occurring 23 ns before the main peak, and the late events extending to 70 ns. The same features appear in Fig. 11.5 taken with an XP2020.

Examples of the mechanisms for early and late events are shown in Fig. 11.6 and explanations for their generation are as follows:

- This is a clean event showing a single anode pulse. All but a few per cent of events are of this type.
- This event is created by elastic backscatter of a photoelectron from d_1 , or any supporting metal structure in the vicinity of d_1 . The backscatter probability is about 0.3 per photoelectron and, for a bialkali layer, laid down on the first dynode, Z_{eff} is 44 (see 5.5). The scattered photoelectron decelerates in the electric field, stops just short of the photocathode, and returns to d_1 after a time of $2 \times t(k-d_1)$. Signals of this kind occur predominantly between 40 and 50 ns in Fig. 11.4.
- A photoelectron may be inelastically scattered from d_1 , carrying away a fraction of the full energy, and a reduced number of true secondaries are also produced. Two anode pulses are thus generated, both undersized. The first pulse may be of sufficient amplitude to trigger the system, in which case the signature is given by (a); otherwise, a late pulse is registered, if the

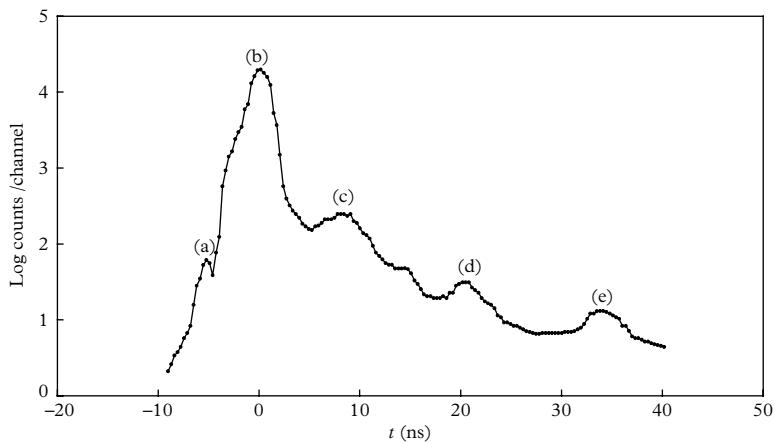


Fig. 11.5. Time response of an XP2020 to a fast flashlamp showing timing (Hungerford and Birch 1996). The early peak, (a), is a prepulse occurring 5 ns before the main peak, (b). Peaks (c), (d) and (e) are late pulses. The transit time quoted by the manufacturer is 28 ns at an HV of 2 kV.

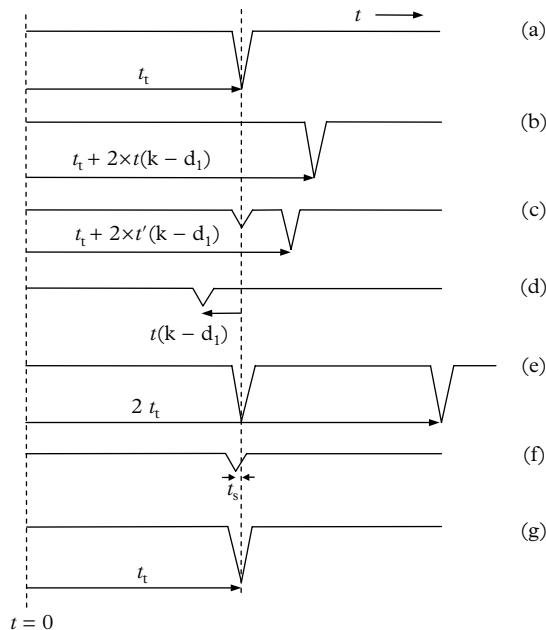


Fig. 11.6. Timing examples of early and late pulses. The cathode to anode transit time is t_D and $0 < t'(k - d_1) < t(k - d_1)$.

- second one is of sufficient amplitude. Clearly, the time taken by the circulating electron is always less than $2 \times t(k-d_1)$, and these events should appear within the time interval of 10 to 40 ns after the main peak, subject to $0 < t'(k-d_1) < t(k-d_1)$.
- (d) A prepulse is generated by a photon incident on d_1 after transmission through the window and photocathode, arriving early by an interval equal to $t(k-d_1)$, ignoring the transit time of a photon from the cathode to the first dynode; it is less than half a nanosecond, even for an 8" hemispherical PMT. Prepulse events are unregistered unless the acceptance trigger threshold is set sufficiently low to allow for an amplitude reduction equal to the gain of the first stage—a factor of 10 to 20. This suggests that an obvious way to suppress prepulses is by raising the trigger threshold. Very roughly, the probability for prepulse events should vary as the ratio of the effective d_1 area to that of the photocathode. Small-diameter PMTs, with a relatively large d_1 area, should therefore show the effect more prominently than the larger ones. A useful diagnostic for confirming the origin of early pulses is to direct the light source at an angle to the photocathode surface so that all transmitted photons miss the first dynode.
 - (e) In this instance, a second pulse is produced at a time equal to the transit time of the PMT. It is known, and easily verified by observation in the dark, that the anode and the last few dynodes glow under electron bombardment by a sufficiently intense signal. There is a finite probability that one of these photons will find a path to the photocathode and start a second cascade. Manufacturers have taken care to shutter the back end to effectively trap these photons. It is thought that anode glow is ultimately responsible for electrical breakdown, brought on by ‘over-volting’. Photon traps are invariably included in linear focussed PMTs in the form of u-shaped enclosures, surrounding the dynode stack. These shields are usually connected to the potential of the middle dynode and also serve as ion traps. They are not incorporated in Venetian blind structures because of interference with photocathode activation; in any case, such dynodes provide inbuilt shuttering. Lubsandorzhiev *et al.* (2006), among others, have verified the occurrence of this phenomenon with a viewing PMT located at the back end, showing that the peaks due to elastic backscatter and afterglow do not coincide exactly. Note that, with the gating mode of Fig. 11.3, the second pulse fails to register in favour of its precursor, which always occurs and takes timing precedence.
 - (f) An early pulse may be generated by high-speed scattering of a photoelectron from d_1 directly to d_2 . The resulting anode signal is undersized, as for (d), and the reduction in transit time is a few nanoseconds.
 - (g) The effects so far described originate from the photocathode and first dynode regions of the PMT, but prepulse effects are also initiated by a

reflex anode—the standard collector in all traditional PMTs (briefly mentioned in 8.2.3). The consequences of an early signal component are particularly severe in $n-\gamma$ pulse shape discrimination (PSD) apparatus, appearing as a discontinuity in the slow–fast plots, and particularly for the higher-energy events (Figs 10 and 11 in Moszyński *et al.* 1991). An early output is observed 1 to 2 ns before the main pulse, because of an induced charge, as electrons from the penultimate dynode approach and pass through the anode mesh, en route to the ultimate dynode. PMTs, other than the fastest types, do not resolve this contribution, but instead generate a pedestal partially merged with the leading edge of the main pulse. The Photonis XP20D0, one of the fastest 2" PMTs available, includes a screened grid disposed between the penultimate dynode and the anode; this reduces the induced current, leading to a reduction in jitter by 20% (Moszyński *et al.* 2006).

Observations (d), (e), and (f) support the earlier statement that the observed detail in a timing distribution depends on the trigger threshold, and on the way in which the delayed coincidence is configured. In the experimental arrangement used by Ranucci *et al.* (1993), the first pulse following the start pulse stops the TAC whereas, for the set-up used by Lubsandorzhev *et al.* (2000), the first pulse following an imposed delay will stop the TDC. Multi-photoelectron pulses may generate a combination of prepulses and late pulses. The probability of recording a prepulse varies approximately as the inverse of the trigger threshold; the rate of such pulses is significantly reduced by increasing both $v(k-d_1)$ and the discriminator level to at least $\frac{1}{4}$ of a photoelectron equivalent. The first event to generate a stop pulse is recorded, but subsequent pulses are ignored. If the requirement is to record all late pulses, such as the second pulse in (e), then the inclusion of a moving gate to the circuit of Fig. 11.3 is necessary. Alternatively, a digital oscilloscope, externally triggered by the CFD signal, would reveal all pulses associated with a main event.

11.4 Detector response function

A fundamental equipment parameter referred to as the detector response function, $D(t')$, is described by Hungerford and Birch (1996). As the terminology suggests, this is a type of calibration that corrects a measured distribution by removing the effects of prepulses and late pulses by mathematical means. It is analogous to subtracting background from signal in photon counting. The detector response function is determined using the same electronics arrangement as intended for measurement, after removing the sample from the instrument and replacing it by a single-photon one.

Fluorescence lifetimes comparable to or even shorter than the instrument response can be uncovered by a mathematical process known as deconvolution by Hungerford and Birch (1996). The resulting time response distribution provides a calibration for the instrument, but it also uncovers timing artefacts in the PMT or, more correctly, invariably attributed to the PMT. The highest amplitude peak, (b), is taken as the timing reference point for all the additional peaks in the spectrum. The width of the main peak of the distribution determines the timing resolution, expressed as a standard deviation, ϵ , or alternatively in terms of the fwhm. The fwhm for the particular light source and PMT combination, illustrated in Fig. 11.5, is ~ 1 ns.

11.5 Afterpulses

11.5.1 Early measurements

Afterpulses are readily observed by triggering an oscilloscope on an LED-induced pulse or a dark pulse. A poor-quality PMT needs only a few minutes of observation to conclude that afterpulses are present, predominantly within 10 μ s of an initiating pulse. They have amplitudes of up to a few photoelectrons, occurring at preferred times with respect to the initiating signal. At the beginning of the 1950s, afterpulses, also referred to as satellite or spurious pulses, posed a serious limitation in the application of PMTs to studies in scintillation spectroscopy, and also in time-of-flight measurements. Among the significant early publications were those of Godfrey *et al.* (1951) and Lanter and Corwin (1952), who attributed the source to positive ions derived from residual gas and from molecules residing on metal surfaces and electrodes. With considerable foresight, the possibility of inter-dynode-generated afterpulses, in addition to those derived from the cathode-to-d₁ region was foreseen. This explanation for the origin of afterpulses is still widely accepted.

A practical consideration in afterpulse studies is that significant research is possible with minimal apparatus. Given a digital oscilloscope, a single-photoelectron source (synchronous or otherwise), and a few hours, one may easily record an afterpulse time spectrum and pulse height distribution of a given PMT by observation of manually repeated single-shot or free-running events. This procedure invariably uncovers unexpected effects that the experimenter should note. If many PMTs are to be examined, it is sensible to automate testing using the knowledge gained by precursory studies.

An example of a typical afterpulse time distribution, measured by the author, is shown in Fig. 11.7. Although this 7" Venetian blind PMT, EMI 9623, is obsolete, its afterpulse performance is, however, representative of more current types. Measurements were made using a 50 ns pulsed green LED of 100 photoelectrons equivalent, feeding a discriminator with a threshold of 0.25 photoelectrons. The afterpulse

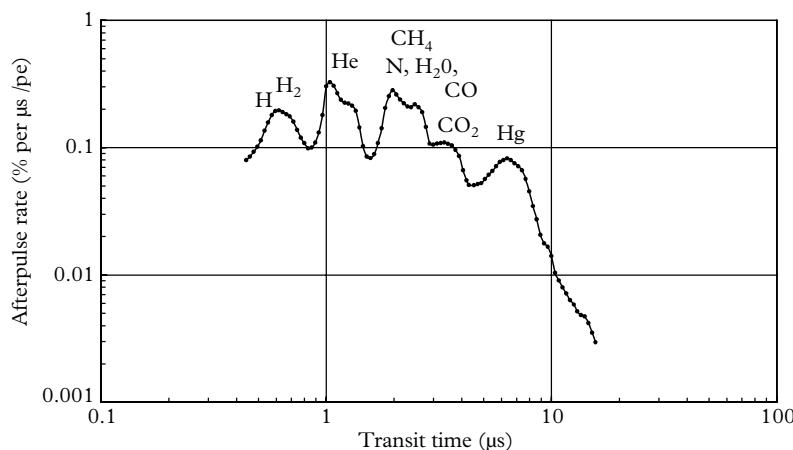


Fig. 11.7. An afterpulse arrival time spectrum of a large-diameter Venetian blind PMT with an S11 photocathode. Data acquisition used the method described in 11.2.2. The annotation above the peaks indicates candidate ions at the times shown. Two shoulders beyond 8 μs may be indicative of antimony/caesium compounds. Note that the peak attributed to helium has a distinct right-hand shoulder similar to that reported by Coates (1973b) for a different tube type. He attributes this shoulder to ions that originate between d_1 and d_2 and yet succeed in reaching the photocathode.

arrival times were measured with a sliding gate of fixed width. Expected arrival times for singly ionized residual gas species are based on the prominent peak at 1 μs , which is known to originate from helium following a short confinement in such an atmosphere. The afterpulse rate for this PMT is about 1% per photoelectron, which is a reasonable figure, considering that helium accounts for $\sim 30\%$ of the total afterpulses. The justification for attributing a peak to mercury stems from the use of a vacuum pump of that type. It should be noted that the EMI 9623B spectrum spans a broader time window compared with the typical 2" fast PMT. This is ascribed to the long k-to- d_1 distance of 15 cm, with correspondingly low field gradients.

Allen (1952) likened afterpulse performance to a ‘crude time-of-flight mass spectrometer’. He succeeded in resealing commercial PMTs on to a gas-handling vacuum system to allow the introduction of known gases. PMT afterpulse spectra were thus calibrated with hydrogen, nitrogen, and argon; Allen was able to identify H_2^+ ; O^+ or N^+ ; O_2^+ , N_2^+ , or CO^+ ; and Hg^+ —all of which have an expected presence in a PMT. Allen’s findings may be summarized as follows:

Afterpulses originate from ions of gaseous impurities and molecules located in the region between the cathode and the first dynode.

They are also produced from a monatomic layer of adsorbed gas on the first dynode (not accepted nowadays).

Some afterpulses originate from inter-dynode ionization, whereby the ion escapes the multiplier structure and terminates on the photocathode.

They occur over an interval of 300 ns to 2.4 μ s in the RCA 5819 PMT.

The time interval between the primary pulse and its afterpulse varies inversely as $v(k-d_1)^{1/2}$. Changing the voltages on the remaining stages has little effect on timing.

These observations agree with the explanations given by other researchers.

11.5.2 Theory of afterpulse generation

The theory that follows is based on the considerations presented by Coates (1973a), and the extension by Incandela *et al.* (1988), assuming the following applies. An afterpulse is generated when a photoelectron, moving from the photocathode to the first dynode, creates an ion from one of the gaseous elements within a PMT. The positively charged ion is attracted to the photocathode; on impact, at time τ later, it creates secondary electrons that are amplified in the usual way to produce an afterpulse. In what follows, we consider single-photoelectron-initiated events only, although dealing with a pulse of photoelectrons should be relatively straightforward. Note two electrons are generated in the process, sharing the original photoelectron kinetic energy. The immediate aim is to predict τ for an ion created within the k -to- d_1 space at a point $s = s_0$. The electron transit time from $k-d_1$ ranges from a few nanoseconds to about 20 ns in large-diameter PMTs, and this is ignored in the analysis, since we already know that τ is of the order of microseconds. Let $V(s)$ in Fig. 11.8 represent the potential at a distance s from the photocathode, measured along the axis of the PMT, and d be the distance between the first dynode and the cathode. The starting point for analysis involves the fundamental equations for the velocity, v , of a charged particle in a potential field, V .

The equation of motion is given by

$$eZ(V - V_0) = \frac{m}{2}(v^2 - v_0^2), \quad (11.1)$$

$$\tau = \int_{s_0}^S ds/v.$$

The time taken to reach a distance $s \leq s_0$ from the cathode follows from (11.1):

$$\tau(s, s_0) = -\left(\frac{m}{2Ze}\right)^{1/2} \int_{s_0}^S \left(V(s_0) - V(s)\right)^{-1/2} ds, \quad (11.2)$$

a frequently quoted formula. Consider a linear field described by

$$V(s) = V(d_1)(s/d).$$

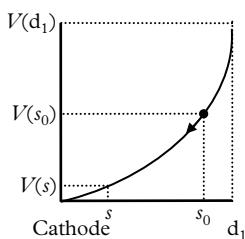


Fig. 11.8. The trajectory of a positive ion, created at $s = s_0$, of charge Ze and mass m , in an electric field $V(s)$.

Then,

$$\begin{aligned}\tau(s, s_0) &= -\left(\frac{md}{2ZeV_d}\right)^{1/2} \int_{s_0}^s (s_0 - s)^{-1/2} ds \\ &= \frac{2md}{(ZeV_d)^{1/2}} (s_0 - s)^{1/2}.\end{aligned}\quad (11.3)$$

The time to reach the photocathode at $s = 0$ is

$$\tau(0, s_0) = \left(\frac{2ms_0d}{ZeV_d}\right)^{1/2}. \quad (11.4)$$

Note, in particular, that $\tau(0, s_0)$ varies as $s_0^{1/2}$ and as $(m/Z)^{1/2}$. Ions created near the first dynode are the last to arrive at the photocathode.

If the electric field varies as the square of s , then $V(s) = V(d_1)(s/d)^2$; $V(s)/V(s_0) = (s/s_0)^2$ and

$$\begin{aligned}\tau(s, s_0) &= -\left(\frac{m}{2ZeV_d}\right)^{1/2} d \int_{s_0}^s [(V(s_0)/V_d)d^2 - s^2]^{-1/2} ds \\ &= \left(\frac{md^2}{2ZeV_d}\right)^{1/2} \sin^{-1} \left(\frac{V(s)}{V(s_0)}\right)^{1/2} \Big|_{s_0}^s \\ &= \left(\frac{md^2}{2ZeV_d}\right)^{1/2} (\pi/2 - \sin^{-1}(s/s_0)).\end{aligned}$$

The time taken to reach the photocathode located at $s = 0$ is therefore

$$\tau(0, s_0) = \frac{\pi}{2} \left(\frac{md^2}{2ZeV_d}\right)^{1/2}. \quad (11.5)$$

Note that transit time is independent of starting position when $V(s)/V(s_0) = (s/s_0)^2$ and therefore we should expect a multi-peaked distribution generated by the mixed gaseous species within a PMT, for such a field distribution.

Potential field plots, representing the k-to-d₁ region, for three fast PMT types follow an $(s/s_0)^2$ distribution quite closely, as illustrated in Fig. 11.9, allowing comparison of (11.5) with measured afterpulse time spectra. The earliest afterpulse is due to H⁺, with $m = 1.67 \times 10^{-27}$ kg, and $Z = 1$; d is 40 mm and we take $V_d = 600$ V with respect to the cathode, and compatible with the operating conditions used by Coates (1973a). Substituting these parameters into (11.5) predicts $\tau(H^+) = 0.185$ μs; transit times for other ions follow from scaling using the $(m/Z)^{1/2}$ dependence given in Table 11.1(b). The close agreement between measurement and predicted transit times in Table 11.1 is gratifying. Assigning the third peak to oxygen, with atomic weight 16, is debatable since it is unlikely to exist

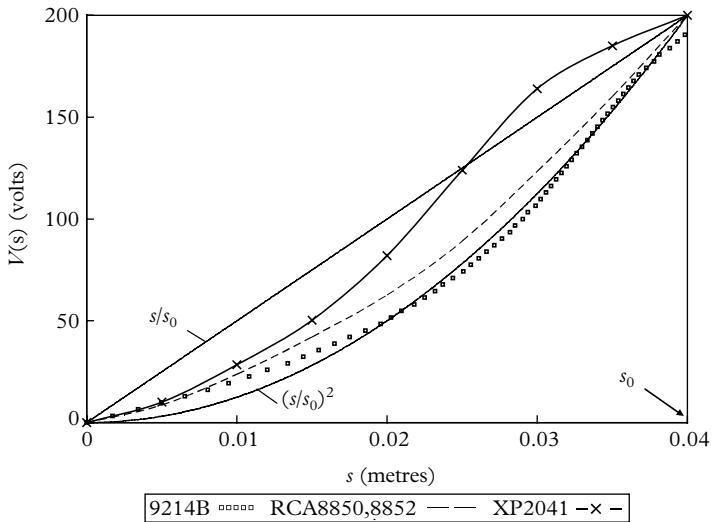


Fig. 11.9. Potential field plots illustrating the approximate square law followed by some fast PMT types. The RCA8850, 8852 curve was derived from Fig. 2 in the article by Coates. The field distribution for the XP2020 is different from these three types because of the inclusion of an accelerator electrode connected to d_7 potential.

Table 11.1 Summary of timing measurements from three authors. All entries are in microseconds. Atomic weights are given in parentheses

(a) Morton *et al.* (1967), RCA7850.

	H_2^+ (2)	N_2^+ (28)	A^+ (40)	Xe^+ (131)
Measured arrival time	0.32	1.17	1.34	2.50
Scaling, \sqrt{m} , based on N_2^+	0.31	1.17	1.40	2.53
Calculated arrival time†	0.38	1.43	1.73	3.14

(b) Coates (1973a, b), RCA8850, 8852.

	H^+ (1)	He^+ (4)	O^+ (16)	—
Measured arrival time	0.187	0.375	0.747	—
Scaling, \sqrt{m} , based on He^+	0.188	0.375	0.750	—
Equation (11.5)	0.185	0.370	0.740	—

(c) Akchurin *et al.* (2007), R7525.

	H^+ (1)	H_2^+ (2)	He^+ (4)	N_2^+ (28)
Measured arrival time	0.170	0.250	0.330	0.650
Scaling, \sqrt{m} , based on He^+	0.165	0.233	0.330	0.620

† Derived by the authors based on knowledge of $V(s)$.

in free form, whereas CH_4^+ and N^+ are known constituents with similar atomic weights of 16 and 14, respectively.

11.5.3 Contributions to understanding afterpulses

A definitive paper by Morton *et al.* (1967) described a handling system for the introduction of known gases, to determine the time signatures of hydrogen, nitrogen, argon, and xenon, introduced separately. All Allen's observations were supported by this work, except the hypothesis that ions may be released at an electrode surface. According to Morton *et al.*: 'the field at any point at the surface of an electrode cannot permit the simultaneous escape of both the secondary electrons of the main pulse and the positive ion producing the afterpulse'. Morton investigated experimental PMTs based on the linear focussed RCA7850 parent type, but incorporating a range of photocathodes. Peaks attributable to molecular H_2^+ and N_2^+ were observed, but without the appearance of H^+ and O^+ . Calculations based on the front-end equipotentials of an RCA7850 are given, indicating only $\sim 10\%$ variation in transit time with the point of initiation on the tube axis. Calculated transit times are averages deduced from an equipotential plot. There is a discrepancy of about 20% between calculated and measured times in Table 11.1(a), but this is not commented upon by the authors. However, given the transit time of N_2^+ , for example, the predicted time for H_2^+ is $\tau = 1.17 \times (2/28)^{1/2} = 0.31 \mu\text{s}$, and similar good agreement applies to argon and xenon. Regarding the size of afterpulses, about four electrons is typical for hydrogen, whereas for nitrogen only a single electron is indicated. Morton *et al.* note that one of the group of the longer-delayed afterpulses disappeared after operating the PMT for a short time at moderate light level. The present author has also noticed this gettering action, particularly for PMTs that have been misused by exposure to light while active, or after the application of excessive HV. These ephemeral afterpulses disappear after a few minutes but always reappear following a break in operation.

Coates (1973a, b) produced a comprehensive study on the RCA 8850 and 8852 PMTs, both of which have GaP first dynodes. Timing distributions exhibit sharp peaks with fine structure pointing to the presence of H^+ and O^+ , contrary to Morton's statement. Coates' flight measurements, shown in Table 11.1, are based on a helium calibration, fortuitously diffused through the envelope of the PMT. An interesting feature of these results is the agreement to within $\frac{1}{2}\%$ between measured and predicted times for H^+ and O^+ . Furthermore, these observations are at variance with Morton's on two counts: H^+ is not evident in the time distribution for the RCA7850 and neither is oxygen. This is possibly because both these elements are highly active chemically. CH_4^+ , which has the same atomic weight as O^+ , is a well-known residual in vacuum devices. Also, N^+ is close in atomic weight to O^+ and, as ion transit times vary as the square root of the atomic weight, this leads to only a 7% difference in transit time making it difficult to distinguish between the two (see (11.2)). The afterpulse-initiating

signal when an ion A is created is represented by $A + e^- \rightarrow A^+ + 2e^-$. Therefore, the initiating signal for an ion created near the photocathode will approach two photoelectrons equivalent and, for ions generated close to d_1 , the initial signal will be close to single photoelectron. This is verified experimentally in Fig. 4 of Coates (1973b).

Akchurin and Kim (2007) investigated the performance of a Hamamatsu R7525, 28 mm diameter PMT. Based on helium calibration, they identify the first prominent peak in the afterpulse time spectrum as H^+ , and a minor shoulder on this peak as H_2^+ . They verified the $\nu(k-d_1)^{-1/2}$ relationship for afterpulse arrival times and showed that the rate is proportional to the input light flux.

Research carried out by Sené *et al.* (1987) on 5" fast PMTs, from various manufacturers, revealed a dependency on the excitation wavelength. The time signatures for 585 nm and for 480 nm (referred to as blue light by the authors), show the same peak positions; however, the spectrum with blue light is a combination of peaks superimposed on a quasi-exponential decay of about 10 μ s duration; this leads to a total rate more than ten times higher than that for 585 nm light. To the present author's knowledge, wavelength dependency has not been reported elsewhere, perhaps because the start of afterpulse generation is invariably assumed to be the release of a photoelectron—the manner of its creation being of no apparent significance. It is customary for investigators to use a monochromatic source, in which case, wavelength dependency goes undetected. It is reported that this component persists, even when gating the photocathode. Sené *et al.* attribute the effect to glass or photocathode excitation. The present author's tentative explanation is that the origin of the effect is located at the first dynode; a relevant investigation would therefore be to angle the light source to avoid illuminating the first dynode. This dynode is known to be photosensitive for the RCA PMT but not for the XP2041 and 9823. Ions created between the first and second dynodes may terminate on the photocathode, producing afterpulses in the generally accepted manner. Ions that terminate on the first dynode, possibly the majority, will produce undersized afterpulses, with a time signature determined by the complicated electron-optics of the d_1 -to- d_2 region.

The work by Sené *et al.* also highlights the importance of the front-end electron-optical design, where afterpulses are concerned. As the 5" RCA8854 is a scaled-up version of the 2" RCA8850 investigated by Coates, the front-end equipotentials of the two tube types therefore have the same $(s/s_0)^2$ form given in Fig. 11.9. This accounts for the distinct peaks evident in the RCA8854. The 9823B and XP2041 have the same complicated potential field distribution, achieved by the inclusion of an accelerator electrode, producing both acceleration and deceleration (see Figs 8.8 and 8.9). The present author suggests that ions created between the cathode and the accelerator are attracted to the photocathode under the influence of the field depicted in Fig. 11.9, while those created between the accelerator and the first dynode terminate on the first dynode and produce undersized afterpulses. The complicated field structure shared by these two PMT types disfavours isochronous timing.

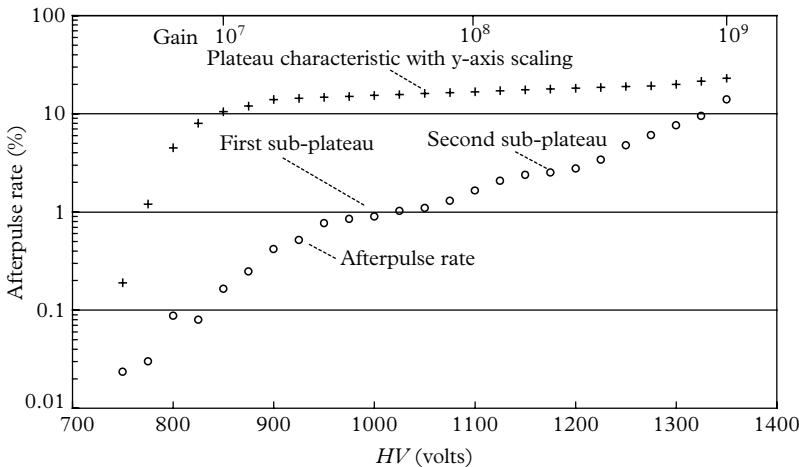


Fig. 11.10. The upper plot is the plateau characteristic, arbitrarily scaled, together with the corresponding afterpulse rate for single-photon excitation of a 28 mm 9107B. Gain and threshold in photoelectrons are reciprocals, with a gain of 10^8 corresponding to a threshold of one-tenth of a photoelectron; the upper abscissa scale highlights the considerable range covered by the photoelectron equivalent scale. Scatter in the afterpulse rate, evident at ~ 800 V, is statistical.

11.5.4 Afterpulses in photon counting

Afterpulse time spectra and rates may be measured with apparatus of the kind illustrated in Figs 11.1, and 11.3. However, in photon-counting applications, the prime concern is the rate of afterpulses. The results shown in Fig. 11.10 were taken with a multichannel scaler (Ortec Easy MCS), with contiguous dwell times set to 20 μ s, thereby including all afterpulses. The memory permits 65,535 measurements, each of 20 μ s duration, with a run time of only 1.31 seconds. The key to this measurement technique is low occupancy; on average, less than one count per dwell time is suitable, when sampling a random light source. There are two ways in which a pair of counts may occur: two random pulses arrive within the dwell time (Type 1), or, one random pulse and its afterpulse arrive within the time window (Type 2). Low occupancy ensures that Type 2 pairs exceed random (Type 1) pairs, leading to efficient extraction of the afterpulse rate, p_a , given in 4.17.1:

$$p_a = \frac{\frac{\text{var}}{\mu} - 1}{3 - \frac{\text{var}}{\mu}}, \quad (11.6)$$

where μ is the computed mean of a run, and var its variance.

Several authors have verified proportionality, down to single-photoelectron excitation, between afterpulse rate and input signal rate. It would be wrong to

assume, in view of this linear dependence, that afterpulses constitute extra signal, and are therefore equivalent to an increase in photocathode sensitivity. The main concern regarding afterpulses is their statistical nature and the time correlation to the initiating signal.

It is revealing to display the afterpulse rate on the same graph as the single-photoelectron plateau characteristic, as in Fig. 11.10. The single-electron peak corresponds to roughly the half plateau count rate, which occurs at about 850 V. Multi-photoelectron afterpulses register in the 750 to 800 V region of the plot, at a rate of about 0.1% per photoelectron. The afterpulse characteristic (\circ) suggests two plateaus: the first between 950 and 1050 V, and the second between 1120 and 1200 V. Note that the number of electrons traversing the d_1 -to- d_2 space is given by the d_1 gain, δ_1 ; this is about 15 for the PMT under study, and the number of afterpulses should increase approximately by this factor over those produced in the k -to- d_1 region. By similar reasoning, the number of afterpulses produced in the d_2 -to- d_3 region will be a factor of $\delta_1\delta_2$ higher, given δ_2 is ~ 5 . It is suggested that the first plateau in the afterpulse rate curve is linked to afterpulses generation in the d_1 -to- d_2 region of the multiplier; such afterpulses are only registered at a higher voltage that compensates for the lost contribution to gain from d_1 . The second plateau refers to the d_2 -to- d_3 region, where two gain stages are ineffective for afterpulse generation. The dead time of the amplifier discriminator is 20 ns in this investigation but it has been shown in 4.19.7 that extending this to 10 μ s removes the contributions from afterpulses by the imposition of dead time.

As discussed in 7.6.2, operating just beyond the knee at an HV of 900 V is the usual recommendation. Historical wisdom in favour of operating in the middle of the plateau characteristic is irrelevant with today's ultra-stable power supplies. The approximate gain scale shown at the top of the figure emphasizes the power law dependency on HV and should discourage operation beyond 1000 V.

11.5.5 Exposure of PMTs to helium

Helium has the property of being able to permeate the types of glass favoured for PMT manufacture. Helium concentration in the atmosphere is sufficient to reduce the lifetime of certain types of PMTs to a few years. Instances of PMTs rendered useless after operating, or simply storing them, in the vicinity of helium dewars have been reported. A publication by Norton (1957), who studied the permeation of helium through nine different glasses, remains the key source of information on this subject. The quantity of gas permeating a plane membrane of material in time t , under steady state conditions, is given by the Fick law:

$$q = K(t)At \frac{\Delta p}{d},$$

where $K(t)$ is the permeation constant at temperature, T ; A is the exposed surface area; Δp is the partial pressure differential across the membrane; and d is the

thickness; $K(T)$ is an important parameter, best determined experimentally. It varies exponentially with temperature as

$$K(T) = K_0 \exp(-Q/RT),$$

where Q is the activation energy, R is the gas constant, and K_0 is a constant of the medium; $K(T)$ ranges over many orders of magnitude for the three most widely used PMT envelopes: it is $\sim 10^{-14}$ for lime soda glass, $\sim 10^{-12}$ for borosilicate glass, and 3×10^{-10} for fused silica. Generally, low K values are associated with higher temperature dependence; K increases by a factor of 10 per $50\text{ }^\circ\text{C}$ for fused silica, 20 for borosilicate, and 50 for lime soda at room temperature. Helium is favoured for calibrating afterpulses as it is easily introduced by placing a PMT (non-operating) into a He-rich atmosphere. There is always a time lag of a day or two between the introduction of helium and the first appearance of afterpulses. This refers to borosilicate but the equivalent time for a quartz window is considerably less. The reader is referred to Incandela *et al.* (1988) and Akchurin *et al.* (2007) to gain an overall appreciation of the theoretical and practical aspects of the subject. Garrard *et al.* (2012) give useful information on the permeability of various glass types used in PMT manufacture.

11.6 Exposure to bright light

11.6.1 Settling time: Medium term

Work of Pettifer and Healey (1974) and Clemesha (1977) concerning transient effects on S20 PMTs are among the few who report on the phenomenon over a time interval of 10 to $600\text{ }\mu\text{s}$; following termination of a relatively bright, pulsed, square-wave signal. The proportion of events in the tail of the distribution, although highly undesirable, is actually small. Afterglow is evident on an oscilloscope by viewing at the lowest millivolt settings. An MCS is ideal for taking measurements of the kind shown in Fig. 11.11. Pettifer and Healey (1974) find that the decay of the enhanced background varies as $\sim t^{-1}$.

11.6.2 Settling time: Long term

PMTs may be exposed to sunlight without apparent harm, although this is not recommended—particularly for infrared III-V PMTs. A degree of exposure is unavoidable at a commissioning stage but it is always advisable to minimize exposure levels. It is important to qualify this statement with the proviso that the HV be disconnected during light exposure. Failure to do so inevitably leads to PMT replacement. Historically, it was usual to wait a day to allow elevated levels of dark current to subside. However, nowadays settling time for the common

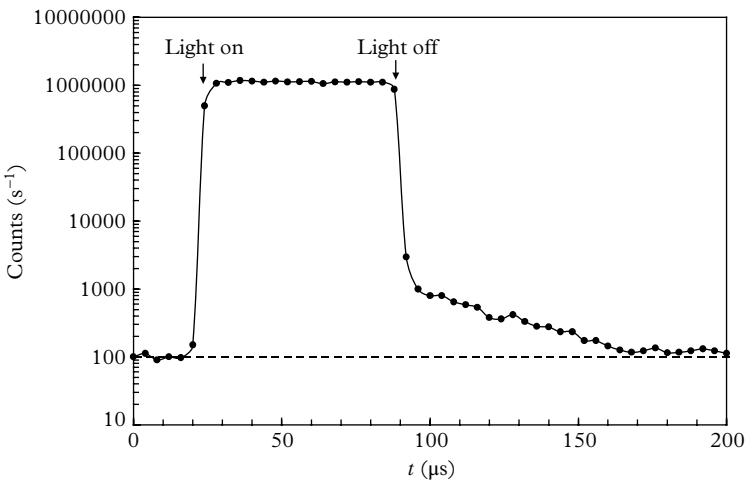


Fig. 11.11. An afterglow curve generated by a signal pulse 70 μs wide. The count rate produced by the source of 10^6 per second, while high, is within the operating range of a PMT. The proportion of events in the tail is exaggerated by the use of a logarithmic scale but it takes more than 100 μs after cessation of the excitation signal to approach the dark current level. A background level of 100 s^{-1} , before exposure, is shown by a dashed line.

PMT types can be measured in minutes. Enhanced dark current consists mainly of single-photoelectron signals, which points to the photocathode as the main source of this type of background. Confirmation follows by open circuiting the photocathode. Certain window materials are known to fluoresce, which may be verified by viewing a dummy window separately exposed. Signal decay is described by an inverse power relationship of the form t^{-n} as opposed to an exponential relationship.

Measurements by J C Barton (private communication) on a selection of PMTs are perhaps the most comprehensive to date. Results in Fig. 11.12 cover the time interval from 10 to ~ 1000 s. Light exposure levels are modest, corresponding to a maximum count rate of 10^5 s^{-1} at 10 s after exposure and, consequently, dead time correction is unnecessary. Figure 11.12 shows the recovery after three different light exposures of 2, 10, and 60 s levels for an S11 9811B. Note that enhancement is approximately linear with exposure time, although the proportionality fails at high intensity. The effect of switching on the HV during exposure is to lift all readings by a factor of about 10. Profiles are similar when a PMT is operated as a photodiode, except the characteristic decay is faster, with $n = 1.27$.

Individual 2" PMT windows exposed to room lighting for several hours and subsequently placed inside a PMT housing (without exposing the PMT itself to any light), produce transient background. The subsequent counting rates, shown in Fig. 11.13, are therefore attributed to glass fluorescence. It is seen that the fluorescence of the high-potassium-content borosilicate glass Soveril 747 gives

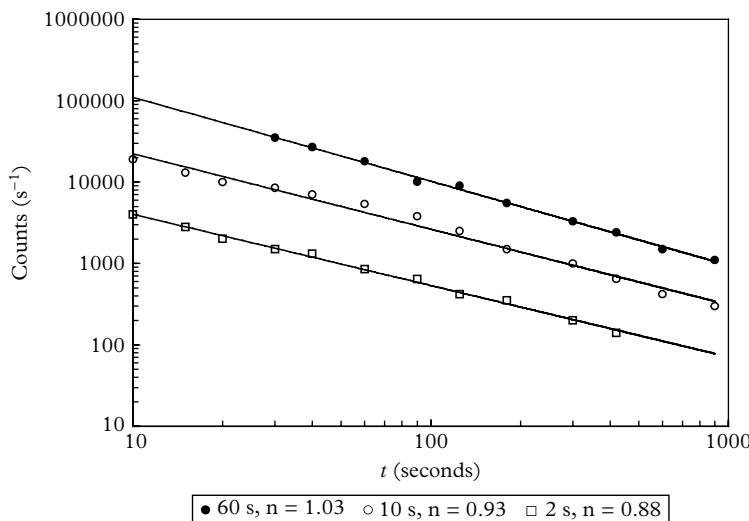


Fig. 11.12. The decay of signal-induced background (residual background subtracted) following the exposure of a PMT to a steady light source for 2, 10, and 60 s in the absence of HV. The first count entered follows 10 s after the cessation of the stimulus. Note the initial count rates are comfortably within the linear capability of the PMT. Signal decay is described by an inverse power relationship of the form t^{-n} .

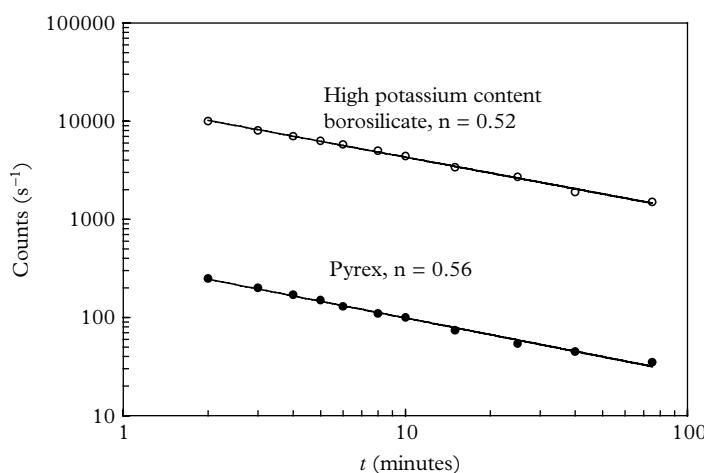


Fig. 11.13. Light emission from separate 2" windows after exposure to room lighting for 24 hours. Note the abscissa units are minutes. A power law relationship follows a t^{-n} law with $n \sim 0.5$.

100 times the light output, compared with Pyrex, for example. Fused silica (quartz), on the other hand, shows little sign of light emission. As with earlier measurements, background decay follows a power law, with the important difference of a $t^{-1/2}$ dependence. The decay rate of this component of light emission is slow compared with the results in Fig. 11.12.

11.7 Summary and conclusions

- All satellite pulses are undesirable, particularly where the signature of the incident light is under investigation. Corrections are possible in certain arrangements (Hungerford and Birch 1996).
- Afterpulses are detrimental in spectroscopy, particularly concerning organic and inorganic scintillators, because of degradation in energy resolution. Their effect in photon counting is twofold: first, in terms of an increased plateau slope and, second, their enhancement of variance at low count rates.
- Afterpulses also originate from ions created between dynodes. Coates (1973a) has shown that afterpulse rates can be halved by judicious tuning of the third and fourth dynode potentials.
- Timing defects appear to be wavelength sensitive and it is important to match any investigations to the actual wavelength of light involved.
- Riley and Wright (1977) showed for systems involving groups of PMTs, in coincidence, that the rate of afterpulses is reduced by an order of magnitude. However, the peaks in the distribution become sharper.
- The use of helium should be avoided for gas flushing: argon and nitrogen have K values in the region of 10^{-15} , making them ideal candidates.
- All PMTs exhibit enhanced background after exposure to high levels of illumination and radiation. In the time range 10 to 10,000 s, the recovery is described by t^{-n} , where $n = 1.1 \pm 0.2$.
- All window materials fluoresce, except fused silica. Light emission follows a much slower decay characterized by $n \approx 0.5$.
- Afterglow is proportional to light exposure for moderate exposure intensities.
- It appears that certain cathode types produce higher levels of transient light than others. The following are ranked in order of level of production, from lowest to highest: Na_2KSb , K_2SbCs , Rb_2SbCs , S11, and Na_2KSbCs . Traditionally, a day's settling time was recommended after exposure to bright light.
- Nowadays, the combination of non-fluorescing window material and developments in photocathode activation has produced PMTs that settle within 10 min, to a level approaching the residual dark current.

- Consequences of long recovery time are undesirable in automated, multi-sample equipment—particularly industrial and diagnostic instruments. Shutting the PMT against ambient light, whenever a new sample is introduced for analysis, is obviously beneficial. Crosstalk, when a weak sample follows an intense one, can be reduced by imposing a delay between sample measurements but this may not be acceptable for throughput reasons. Selecting the appropriate PMT may help. LIDAR and fluorescence studies are also affected, even if gating is used.

It is telling that, some 75 years later, most of these early observations on transient behaviour remain valid but our present understanding is still incomplete and sometimes contradictory. However, PMTs of current manufacture have lower rates of satellite signals than their predecessors. Modern designs include a high degree of shuttering to frustrate light feedback. Ion traps and other means, combined with electron-optical design, also contribute to low afterpulsing.

REFERENCES

- Akchurin, N. and Kim, H. (2007). A study of ion initiated photomultiplier after-pulses. *Nucl. Instr. and Meth. in Phys. Res. A*, **574**, 121–6.
- Allen, J. S. (1952). *After-pulses in Photomultiplier Tubes. Los Alamos Report, LA-1459*. Los Alamos Scientific Laboratory of the University of California, Los Alamos, NM.
- Baldini, A. *et al.* (1996). The photomultiplier test facility for the reactor neutrino oscillation experiment CHOOZ and the measurement of 250 8-in. EMI 9356KA B53 photomultipliers. *Nucl. Instr. and Meth., in Phys. Res. A*, **372**, 207–21.
- Bollinger, L. M. and Thomas, G. E. (1961). Measurement of the time dependence of scintillation intensity by a delayed coincidence method. *Rev. Sci. Instr.*, **32**, 1044–50.
- Bonitz, M., Meiling, W., and Stary, F. (1964). Prepulses in photomultipliers. *Nucl. Instr. and Meth.*, **29**, 314–17.
- Clemesha, B. R. (1977). Transient noise in photomultiplier tubes. *J. Phys. E: Sci. Instr.*, **10**, 814–16.
- Coates, P. B. (1973a). The origins of afterpulses in photomultipliers. *J. Phys. D: Appl. Phys.*, **6**, 1159–66.
- Coates, P. B. (1973b). A theory of afterpulse formation in photomultipliers and the prepulse height information. *J. Phys. D: Appl. Phys.*, **6**, 1862–9.
- Garrard, C. C., McAlpine, R. M. and Wright, A. G. (2012). *The Effect of Exposure to Helium on Photomultiplier Performance and Lifetime*. ET Enterprises Reprint Series R/P101. ET Enterprises Ltd, Uxbridge. Available at http://www.et-enterprises.com/files/file/technical-information/rp101_effect%20of%20exposure%20to%20helium%20on%20photomultiplier%20performance%20and%20life_time.pdf (accessed 3 Feb 2017).

- Godfrey, T. N. K., Harrison, F. B., and Keuffel, J. W. (1951). Satellite pulses from photomultipliers. *Phys. Rev. Lett.*, **84**, 1248–9.
- Hungerford, G. and Birch, D. J. S. (1996). Single-photon timing detectors for fluorescence lifetime spectroscopy. *Meas. Sci. Tech.*, **7**, 121–35.
- Ianni, A. *et al.* (2005). The measurement of 2200 ETL9351 type photomultipliers for the Borexino experiment with the photomultiplier testing facility at LNGS. *Nucl. Instr. and Meth. in Phys. Res. A*, **537**, 683–97.
- Incandela, J. R. *et al.* (1988). The performance of photomultipliers exposed to helium. *Nucl. Instr. and Meth. in Phys. Res. A*, **269**, 237–45.
- Lanter, R. J. and Corwin, R. W. (1952). Spurious pulses from the Type 5819 photomultiplier tube. *Rev. Sci. Instr.*, **23**, 507–8.
- Lubsandorzhiev, B. K. *et al.* (2000). Studies of prepulses and late pulses in the 8" Electron Tubes series of photomultipliers. *Nucl. Instr. and Meth. in Phys. Res. A*, **442**, 452–8.
- Lubsandorzhiev, B. K. *et al.* (2006). Photoelectron backscattering in vacuum phototubes. *Nucl. Instr. and Meth. in Phys. Res. A*, **567**, 12–16.
- Morton, G. A., Smith, H. M., and Wasserman, R. (1967). Afterpulses in photomultipliers. *IEEE Trans. NS*, **14**, 443–8.
- Moszyński, M. *et al.* (1991). Comparative study of new 130 mm diameter fast photomultipliers for neutron detectors. *Nucl. Instr. and Meth. in Phys. Res. A*, **307**, 97–109.
- Moszyński, M. *et al.* (2006). New Photonis XP20D0 photomultiplier for fast timing in nuclear medicine. *Nucl. Instr. and Meth. in Phys. Res. A*, **567**, 31–5.
- Norton, F. J. (1957). Permeation of gases through solids. *J. Appl. Phys.*, **28**, 34–9.
- Pettifer, R. E. W. and Healey, P. G. (1974). Signal induced noise in a 56 TUP photomultiplier. *J. Phys. E: Sci. Instr.*, **7**, 617–20.
- Ranucci, G. *et al.* (1993). Performance of the photomultiplier EMI 9351 for underground physics applications. *Nucl. Instr. and Meth. in Phys. Res. A*, **333**, 553–9.
- Riley, R. J. and Wright, A. G. (1977). The effect of photomultiplier afterpulses in coincident systems. *J. Phys. E: Sci. Instr.*, **10**, 873–4.
- Sené, M. R., Hawkes, N. P., and Finlay, D. J. S. (1987). Afterpulsing in 130 mm diameter photomultiplier tubes. *Nucl. Instr. and Meth. in Phys. Res. A*, **254**, 591–9.
- Stevens, S. S. and Longworth, J. W. (1972). Late output pulses from fast photomultipliers. *IEEE Trans. NS*, **19**, 356–9.

12

Environmental considerations

12.1 Introduction

Environmental conditions affect PMT performance. Consideration has already been given to some relevant topics, such as interfacing to small- and large-area scintillators (3.4 to 3.10); exposure to bright light (11.6); the temperature dependence of background (6.6.3, 2.9); the temperature dependence of cathode resistivity and QE (2.7, 2.8); ionizing radiation and gamma background (2.10.2, 6.6.4, and 6.7); and the effect of cosmic rays (6.6.4, (see also Appendix B)). We expand on these in the sections that follow, in addition to considering further ones.

12.1.1 Performance in weak magnetic fields

There are two aspects to this topic. One concerns how standard product functions in the presence of modest magnetic fields of strength comparable with the field of the earth. The other consideration focuses on devices capable of functioning in the presence of high magnetic fields: those that are four orders of magnitude in excess of the earth's field.

Undoubtedly, the foremost requirement is that of shielding against modest magnetic fields, such as the field of the earth (~ 1 gauss or 10^{-4} tesla). Although this field is essentially constant over the surface of the earth, changes in sensitivity arise when the orientation of equipment is changed—such as for portable instrumentation, for example. Vacuum pumps and electromagnetic relays are also sources of interference but are usually transient in nature. Actual performance in a magnetic environment depends on the following considerations:

- Short path lengths are least affected by magnetic fields and, consequently, small-diameter variants perform better than those of larger diameter, particularly hemispherical PMTs
- Fields directed along the axis of end window PMTs produce the smallest effect. With box-and-grid, linear focussed, and Venetian blind types, the output is halved for an applied field of about 2×10^{-4} tesla, directed along or across the dynodes. For the metal-channel miniature PMTs, 5600 series, the comparative figure is 10^{-2} tesla. An example of the benefit of a magnetic shield is shown in Fig. 12.1.

Interference caused by magnetic fields can be reduced by covering a PMT with a wrapped shield of MuMetalTM[†]. A basic but sufficient solution may be achieved by wrapping a 0.125-mm thick sheet of material around the cylindrical section of the PMT. Further turns may be applied to achieve a higher level of shielding. Alternatively, a formed shield in the shape of a thin-walled cylinder may be employed, typically 0.5 mm thick, to achieve an even higher degree of shielding. Manufacturers offer contoured shields for PMTs with a window diameter

[†] Registered trademark, Magnetic Shield Corporation.

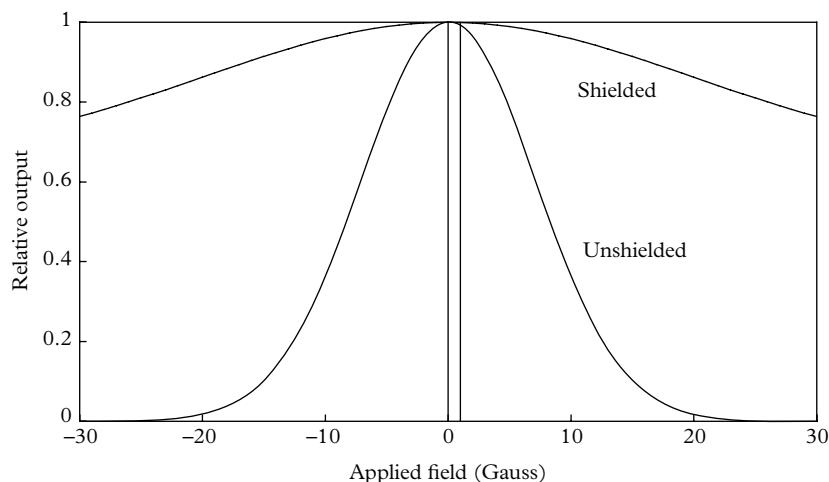


Fig. 12.1. A wrapped MuMetal™ shield of thickness 0.125 mm reduces the sensitivity to a magnetic field directed across the dynodes. Manufacturers provide curves such as this one for all products. The field of the earth is approximately 1 gauss, as shown.

exceeding 50 mm. Ideally, to be most effective, a cylindrical shield should extend beyond the PMT window, but this invariably interferes with the optical interface and it is seldom practicable.

A magnetic shield also acts as an electrostatic one but it does not provide intimate contact with a glass envelope, as may be achieved with either Aquadag™[‡] (Acheson Industries, ICI) or a thin, conductive, and self-adhesive copper sheet, for example. An electrostatic shield, in intimate contact with a glass envelope, should be connected to cathode potential and, if sufficiently insulated, the magnetic shielding may be earthed. Common practice is to connect the magnetic shield via a $10\text{ M}\Omega$ safety resistor to the cathode potential, sometimes with a capacitor in parallel. However, unsatisfactory behaviour with these configurations has been reported by Bristow *et al.*(1995), who found that gain stability, when gating, was compromised by the presence of a ‘safety’ resistor. This effect may be related to high capacitance, of the order of 100 pF, formed by the shield, the glass envelope, and the cathode. Direct connection to cathode potential provided the solution in this instance. Achieving a satisfactory level of shielding, without incurring noise and instability problems, may require experimentation. Fortunately, there is a finite number of combinations for insulation and grounding, at least one of which should prove satisfactory.

[‡] Registered trademark of ICI Ltd.

12.1.2 Performance in high magnetic fields

The emphasis so far is on ways to reduce the effect of magnetic fields on the performance of standard PMTs. A challenge lies in providing PMTs that can operate within the high magnetic field environment demanded in certain high-energy physics experiments. This has to be accomplished without shielding because of interference with the magnetic field distributions within the detector. There are two types in the family of low-gain devices. Triodes and tetrodes (R2148 and R2149) provide gain of about 6 and 16, respectively, at a flux density of 1 tesla. High-gain devices such as the R6504, with 19 fine-mesh dynodes, have gain of 2×10^5 . All PMTs suitable for operation in high magnetic fields show angular dependence and loss in sensitivity at 1 tesla.

12.1.3 The effect of electric fields

Ideally, the glass envelope of a PMT should be insulating over its entire interior surface. However, in the course of photoactivation, alkali metals settle on the walls, which is both undesirable and unavoidable. Some PMTs show patches of photocathode material through unwanted antimony deposition on the walls and on internal shields; this is readily visible in Venetian blind multipliers, which do not include internal shields, but other multiplier structures are affected to a lesser extent. In the extreme, the outcome under operation is one of unstable light emission and changing electrical potentials, first reported by Krall (1967). One way to improve on this is to maintain the envelope at a fixed potential. The discussion so far refers to the use of conventional photoactivation techniques. PMTs manufactured by the transfer technique should be free from these defects. The PMT window is particularly sensitive to electric field gradients, and the photocathode may be permanently damaged through poisoning by the migration of sodium ions (Lavoie 1967).

PMTs may be operated with either positive or negative HV. Positive voltage is applied at the anode, with the photocathode at ground potential, or negative voltage is applied at the photocathode, with the anode at ground potential. However, the quality of performance is not always the same for these two modes of operation. To understand the reasons for this, it is necessary to investigate PMT performance in the presence of electric field gradients generated by the proximity of an associated metal enclosure—an unavoidable adjunct with all PMTs. Major manufacturers carried out these investigations many years ago, reaching conclusions that are valid today. The glass tubing and window are a source of gain instability and excessive background, unless certain rules are followed. This behaviour may be investigated by performing experiments with the arrangements shown in Fig. 12.2. A conductive coating applied to the cylindrical part of the envelope, but drawn back from the window, is illustrated in Fig. 12.2(a). Thin copper foil with conductive adhesive backing is ideal for this purpose. Anode signal current and dark current vary in the way depicted in Fig. 12.3, as a function of V .

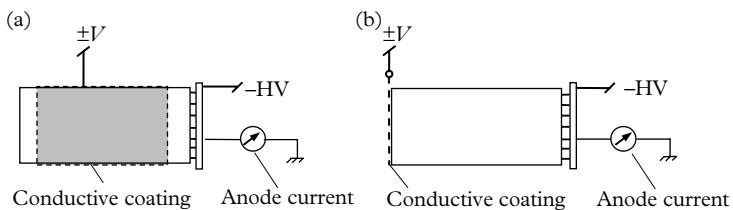


Fig. 12.2. Measurement schemes that lead to better understanding of PMT behaviour.

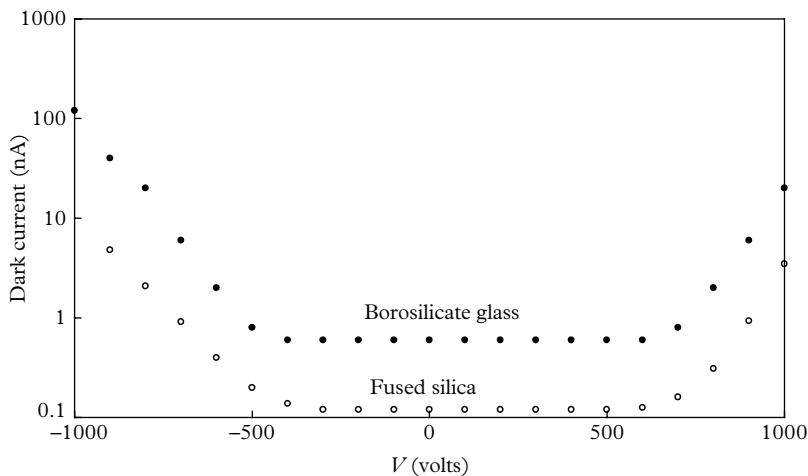


Fig. 12.3. The effect of applying a potential, V , to the envelope of a PMT, using the arrangement shown in Fig. 12.2.

Further investigations with the voltage, V , applied to the window, or to a metal ring on the periphery of the window, show similar effects. Optical shutters, if located too close to the PMT window, generally cause instability in the same way, and electrical interference if magnetically operated. The photocathode of a side window PMT is remote from the envelope and yet it shows the same effects. Investigations indicate that the statement ‘anything in contact with the PMT window must be connected to cathode potential’ can be relaxed to ‘near cathode potential’, although this would appear to have no practical advantage. A related test is to surround a bare PMT with a conducting sleeve of variable diameter. Kume *et al.* (1994) find that dark current decreases, by a factor of 3, and subsequently stabilizes once the gap exceeds 4 mm. The critical distance of 4 mm is also reported in the Philips (1994) handbook, although in this case the gap was filled with insulator. One may speculate that quartz performs better than borosilicate glass because of its higher purity. Also, it may be significant that the window and the first few centimetres of the leading section of a ‘quartz’ envelope are both of fused silica—a highly purified insulating material.

The challenge is essentially one of how best to deploy a semi-insulated HV device, such as a PMT, within a grounded confinement. It should be clear from this discussion that operating with the photocathode grounded is the more stable of the two possible polarity modes. Background is generally lower than can be attained with negative HV. Experimentation, previously discussed, confirms the 4 mm rule for a window and for a bare envelope when negative HV applies. We can circumvent this rule by covering part of the envelope with conducting paint or tape; for safety reasons, protection should include an insulating sleeve. Aquadag™, a graphite-based coating, is low cost but sufficiently conductive to maintain a constant potential along the entire length of the glass envelope. It is customary to extend the coating to within a few centimetres of the pins, with a painted tail in contact with the cathode pin. This material has a tendency to rub off, and any traces of it in unintended places must be avoided. Thin copper foil, backed with conductive adhesive, is easy to apply and it has considerably lower resistivity than Aquadag. A thin metal finger tucked under, and connected to the foil, provides a permanent connection to the cathode pin. Both shielding configurations need a protective and insulating cover, for which a thermoplastic extruded tube (heat shrink) is ideal. Care is necessary whenever heat shrink sleeving is used to cover components of slightly different diameter. The unavoidable differential tension will ultimately cause movement of the sleeve towards the narrowed part. It has been known to creep over the back end of a PMT, finally coming to rest against the pins. Some users, as a safety measure, make contact between a shield and the cathode pin via a $10\text{ M}\Omega$ resistor. However, this resistor may burn out should the shield touch ground potential, leaving the system in an undefined state. Figure 12.4 shows some of the available options for screening: all schemes are intended for negative polarity but may be used with positive polarity. The conducting coating stops at least 1 mm short of the window. They are also suited to positive operation, in which case the conducting material and insulation may be

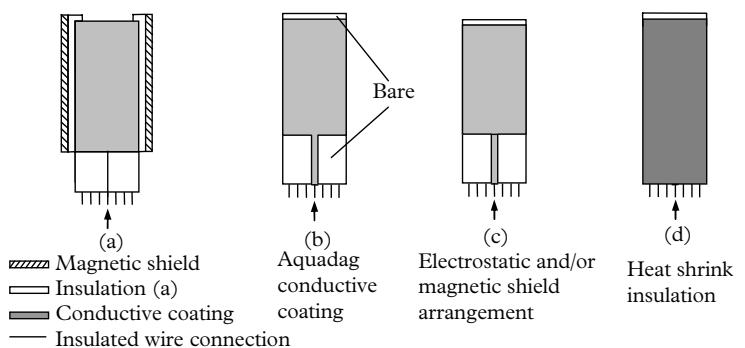


Fig. 12.4. Various ways to ‘dress’ a PMT for operation under negative HV. The cathode pin is indicated by an arrow. The scheme in (a) has the disadvantage of window encroachment and a significant increase in the housing diameter.

taken flush with the window or slightly pulled back as not to interfere with an attachment to the window, such as a crystal or light guide, for example.

12.1.4 PMT enclosures, modules, and housings

The lack of agreed nomenclature is the first obstacle to overcome; even manufacturers are inconsistent in their use of terminology, so it is important to understand what is actually intended. The basic form of a conventional PMT is a cylindrical glass envelope with pin connections—this is referred to as ‘bare’. PMTs are frequently used in this manner, especially when operation is with the photocathode grounded. A PMT may be configured for use with negative HV, but success with this mode of operation may depend on how the PMT is dressed, and attention to finer points is covered in the sections that follow. An enclosure comprises an addition to a bare PMT, such as the inclusion of an insulating thermoplastic sleeve, or a conductive coating of a carbon suspension such as Aquadag™, itself covered with an insulating sleeve. Further options are a fibre-glass sleeve suitable for ambient and high-temperature operation, or a metal shield—magnetic or otherwise. Some of these options are shown in Fig. 12.4. PMTs supplied in a fibre-glass or metal sleeve are generally potted and terminated with a plug, allowing exit of HV and signal cables, and spring loading if required. These are described as modules but they are also referred to as packages or assemblies, in the technical literature. They have gained wide acceptance in academia and in industry as cost-effective means for producing a light detector. An example of a detector module is shown in Fig. 12.5; more advanced modules may include an HV power supply. A significant advantage is that HVs are confined within the enclosure; internal HV supplies operate with 5 or 12 V input. The option of a mounting flange, located at the front of a module, provides a convenient means of interfacing to analytical instrumentation; light-tightness is assured by including ‘O’

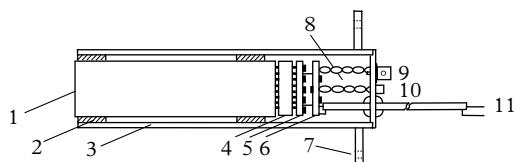


Fig. 12.5. Typical light detector module with a PMT dressed as in Fig. 12.4(b) or (c); heat shrink sleeving, as in Fig. 12.4(d), may be included. A compact Cockcroft–Walton HV supply may be added by extending the length of the module. This design is suitable for positive or negative HV. Legend: 1, photocathode; 2, insulation (RTV or PTFE, depending on PMT dressing); 3, magnetic or fibre-glass cylinder; 4, socket; 5, voltage divider; 6, preamplifier or discriminator; 7, flange with mounting holes positioned as required; 8, this space may be potted or packed with soft foam to aid light-tightness; 9, signal output via twisted pair, BNC connector, or short coaxial line; 10, HV via grommet in back plate; 11, low voltage input.

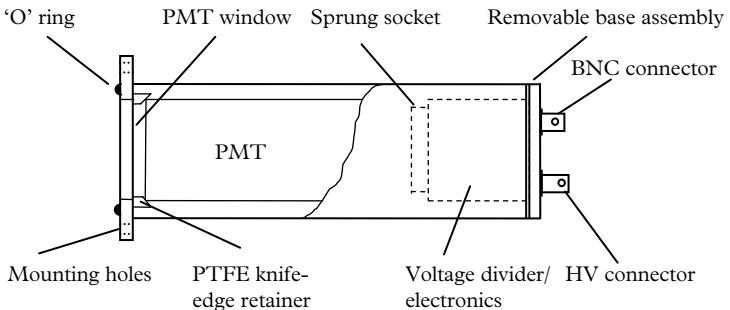


Fig. 12.6. An ambient housing for 51 mm diameter PMTs, with options of a flange and a built-in magnetic shield.

rings. Modules may be rear mounted by the inclusion of a flange or by making use of the PMT socket option. The market for small assemblies is exclusively served by Hamamatsu, who draw on their range of ultra-small side window and end window PMTs, such as the Metal Package 7600 series. These modules are generally boxed, whereas the 25 and 30 mm versions, also available from ET Enterprises, are cylindrical.

'Housings', such as the one depicted in Fig. 12.6, are the natural choice for a one-off system, for instrumentation under development, or where the annual requirements are small. They are available for side window, 30 mm, and 51 mm end-on PMTs. Cooled housings, primarily for use with GaAs(Cs) and multi-alkali photocathodes, are also available. Important considerations include using a centralizing ring located on the rim of the window. The use of negative HV requires a quality insulator, such as PTFE, with chamfered knife-edge cross section. Polyoxymethylene (POM), a thermoplastic synthetic polymer also known as acetal and polyformaldehyde, is unsuitable for this purpose. It is opaque and white in its native form but DelrinTM (DuPont), for example, is black. This material is easily machined with an attractive finish and is suitable for use in other parts of the housing, away from the PMT window. Cleaning the window and the centralizing ring with isopropyl alcohol is recommended, prior to fitting into a housing. Where PMTs are intended for low-light-level DC detection, it is especially advisable to clean the envelope, the pin connections, and the socket—although few users make the effort.

Side window and end window housings are offered by only two manufacturers. If fitted with appropriate adaptors, such housing can accommodate all end window PMTs with a diameter of up to 51 mm. Additional features of all housings, whether cooled or not, are gating and other electronic circuits that may be included; performance with different voltage dividers may be investigated, and the PMT is easily removed or exchanged. A recommendation concerning all housings and modules is to minimize heat generation from the voltage divider. Those that run hot tend to compromise reliability. In this situation, the user should

consider adopting an active or semi-active divider (see Chapter 13). This recommendation is particularly relevant to cooled housings, for obvious reasons.

There are two precautions to observe when operating a cooled housing. First, the PMT should only be removed at room temperature, to avoid producing condensation within the chamber. Second, the specifications quote a figure for ΔT , the maximum differential that can be achieved with respect to ambient temperature; ΔT is typically 50 °C but note temperature regulation ceases under ‘flat-out’ operation (the preferred default setting for the unwary), following any change in room temperature. The optical interface to a cooled housing is poor compared with that attainable with a module or bare PMT. This is because of the inclusion of a double-glazed and evacuated windows that reduces heat conduction into the chamber. There is a gap between the double window and the PMT faceplate to facilitate cooling. The total distance between the outside window of a cooler and a PMT window is of the order of 30 mm. This is an undesirable feature in terms of loss in solid angle, making the standard PMT cooler best suited to collimated or focussed light only. Reflection losses incurred through four additional air–glass interfaces total about 16 %. Attempts to design a cooled housing are unlikely to be cost effective. The usual outcome is a damaged PMT—a consequence of electrical breakdown or uneven cooling. According to the manufacturers, the latter causes migration of caesium to the photocathode, thus changing its characteristics.

12.1.5 Housings for scintillators

The type of housing preferred by manufacturers of scintillation crystals is illustrated in Fig. 12.7 for a 51 mm diameter PMT operated at positive HV. Mounting and interfacing details for the crystal and PMT have been omitted, but they can be found in the manufacturers’ literature. The back end is the relevant part of the housing for the present discussion because it illustrates an effective way of producing a light tight enclosure for this and other applications. The selected PMT is one with flying leads that mate to a loose fitting cap, positioned with respect to the housing, as shown. The length of the assembly can be set precisely and independent of the PMT dimensional tolerances. Once correctly adjusted, a seal is made with RTV or similar material, and the flying leads are trimmed and soldered. Note any pins not carrying a lead must also be soldered to avoid light leaks. Not only does this configuration provide a light tight and hermetic seal but

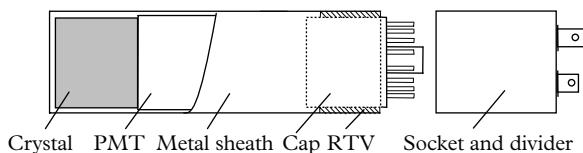


Fig. 12.7. An integral in-line assembly offered by scintillator manufacturers for use with positive HV. The crystal is bonded to the PMT window.

external access is available to the pins. A wired socket carrying HV and signal leads is the simplest arrangement, but a plug-in divider assembly with a preamplifier, for example, may be used in conjunction with the unit. Note that the crystal is bare and permanently attached to the PMT window in this configuration but some users prefer to use a detachable crystal encapsulated within a metal can. This is quite common but it is imperative to connect the can to the potential of the photocathode for the reasons outlined in 12.1.3.

Black insulating tape may be used as part of the fixing arrangement, for example, locating PMTs in large-volume plastic scintillation counters. A capped PMT is introduced into a thin sleeve of slightly larger diameter and taped in the following way: first, the diameter of the cap flush with the outside surface of the sleeve is built up by using several turns of tape; second, several turns of a wider tape is used to cover both the sleeve and the tape applied in the first step. This method overcomes the problem of tape migration by avoiding differential tension previously mentioned. There is a similar problem with heat shrink material if cut too close to the back end of the envelope—in time, the socket may be forced off the base.

12.1.6 Light leaks and electric fields

PMTs must be shielded against ambient light—a statement that at first would seem to state the obvious. However, since we are dealing with devices that are capable of detecting single photons, no light other than signal must reach the photocathode. This is more challenging than is often assumed, particularly when building prototype housings: telltale signs are hurriedly applied black paint, tape, RTV, and, in desperation, chewing gum. External light can easily enter a poorly designed housing through screw heads, microscopic holes and gaps in the enclosure, or even through disconnected bulk head terminations, such as a BNC and an SHV. The presence of externally generated light is easily detected, for example, by switching off room lighting or temporarily covering the equipment with a black cloth. A torch is useful for locating the actual position of a light leak. This experience, for a first-time user, is usually met with astonishment but underlines the sensitivity of PMTs.

Uncovering sources of internally generated light in housings is not straightforward because visual investigation is impossible. However, PMT manufacturers have learned from years of operational experience, gained by themselves and by customers, where these sources lie. Many insulators admit light under a voltage gradient, and any in the line of sight of the photocathode need total elimination. It is customary to locate the PMT window in a metal or insulating ring, as shown in Fig. 12.6; the adoption of high-quality insulating material is mandatory where negative HV is used and, furthermore, only PTFE cleaned in isopropyl alcohol has been found entirely satisfactory. As discussed in 11.3, secondary electrons created within a PMT generate light by ionization and excitation, particularly in the anode region of the multiplier, and a small fraction may be detected by the photocathode.

12.2 Operation in harsh environments

Demand for products by the aerospace industries in the 1960s (artificial satellites) and the oil exploration companies (oil well logging), called for the development of rugged PMTs. In space exploration, PMTs are generally inactive during launch, but must survive undamaged, while, in oil well logging, measurements while drilling (MWD) are taken at high temperature in the presence of severe levels of shock and vibration. Satellite-borne missions may involve PMTs in a diverse range of payloads, whereas oil well logging is predominantly concerned with sodium iodide scintillation counters, but with growing interest in $\text{LaBr}_3(\text{Ce})$ and related scintillators. These extreme environmental conditions call for special PMTs and additional tests beyond catalogue specifications. There are three temperature-sensitive considerations for a rugged NaI(Tl) detector, none of which is easily quantified because of challenging experimental conditions:

- NaI(Tl) light output drops as a function of temperature: measurements by Rozsa *et al.* (1990) indicate a drop in output of the order of 50 % between 20 and 200 °C.
- Photocathode sensitivity for blue light is 40 % that at room temperature (Persyk *et al.* 1976), implying that the combined deterioration will cause a drop in sensitivity to 20 % that at room temperature.
- Dark current limits the performance of PMTs, principally at temperatures exceeding 100 °C.

12.2.1 Shock and vibration

Shock and vibration tests are meaningful provided that the PMT mounting is clearly specified for the test. If the PMT is to be housed, tests are best performed in the housing. In some instances, as an added complication, the specification includes operation with the PMT active. Commercial shock and vibration test houses are usually prepared to offer their expertise in the co-development of a suitable housing. Too often, the designs actually couple mechanical energy to the PMT, rather than absorbing it. A common mistake lies in using insufficiently rigid potting material. It is customary to specify a maximum rating and a maximum recommended level as follows: MR is the maximum design level that must not be exceeded; maximum recommended levels (MRL) refer to the permitted continuous level of operation. There may be a difference of a factor of 2 between these two specifications. The following standard shock and vibration specifications are offered by the industry: IEC 60068, JISC0040, JISC0041, MILSTD 202F, and MILSTD 810E.

Shock tests are performed under gravity with a deceleration profile of a half sine wave and for an agreed duration, typically within the range of 0.5 to 2.5 ms.

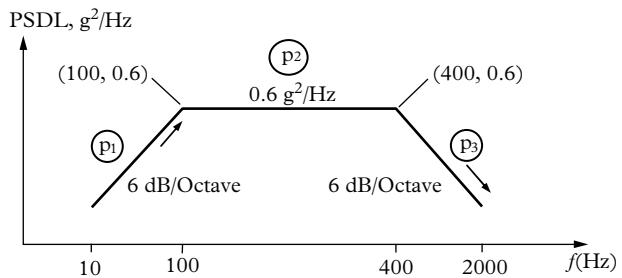


Fig. 12.8. An example of a random-vibration PSDL profile. Also referred to as an acceleration spectral density level (ASDL) profile, to stress the acceleration aspect.

Shocks are applied in each direction for a specified acceleration of up to $1000\hat{g}$, and for a specified number of repetitions—typically 30 to 60.

Sinusoidal vibration is relatively easy to perform and is suitable for less demanding applications. The frequency range is typically 20 to 2000 Hz at $30\hat{g}$ constant acceleration. This test is useful to manufacturers for identifying microphony caused by resonances in meshes and wires within a PMT; the HV must be applied for a meaningful test but it may be reverse biased for ease of measurement. An alternative method available to manufacturers is to use an evacuated but unprocessed PMT for this purpose. According to McAlpine and Wright (1994), resonances occur within 1 to 2 kHz and are visible to the human eye. A design adopting short internal leads obviously withstands shock and vibration, better than a long PMT.

A more exacting and relevant test is based on random vibration, specified in terms of power spectral density level (PSDL) over a bandwidth Δf , with: $\text{PSDL} = g_{\text{rms}}^2 / \Delta f$. Figure 12.8 is an example of a PSDL density profile, as a function of frequency, with a roll-on-and-roll-off rate of power application.

The equivalent g_{rms} follows from the integral of the profile: the equation of the roll-on is $y = (f/a)^2$, with $f = 100$, and $y = 0.6$, leading to $a = 129$. The power in the roll-on is

$$p_1^2 = \int_{10}^{100} (f/129)^2 df = 20.$$

The roll-off equation is $y = (a/f)^2$, with $f = 400$, and $y = 0.6$, yielding $a = 310$. The power in the roll-off is

$$p_3^2 = \int_{400}^{2000} (310/f)^2 df = 192.$$

The contribution from the flat sector is

$$p_2^2 = 0.6 \times 300 = 180.$$

The composite power is

$$P = (p_1^2 + p_2^2 + p_3^2)^{1/2} = 20g_{\text{rms}}.$$

12.2.2 Operation at high temperature

The hot bialkali photocathode (Na_2KSb) can survive and function at temperatures of 200°C , although the dark current may exceed $1\ \mu\text{A}$ under such conditions. The rapid rise in dark current with temperature is illustrated in Fig. 12.9.

High dark current and temporary loss of photocathode sensitivity combine to reduce the quality of performance, particularly resolution, which degrades from 12 % at room temperature to 15 % at 160°C for the PMT illustrated. A typical detector response is shown in Fig. 12.10. The shift in the ^{137}Cs peak to lower channels indicates a reduction in QE, with a smaller contribution from gain reduction. A loss in resolution is strongly affected by shot noise associated with the dark current, rather than by the dark current itself. Although the two aspects of signal are related, this observation points to the importance of reducing the bandwidth of associated electronics to improve resolution at high temperature.

The loss in resolution is caused by shot noise associated with dark current at high temperature. This manifests itself as a steep pedestal, referred to as a noise

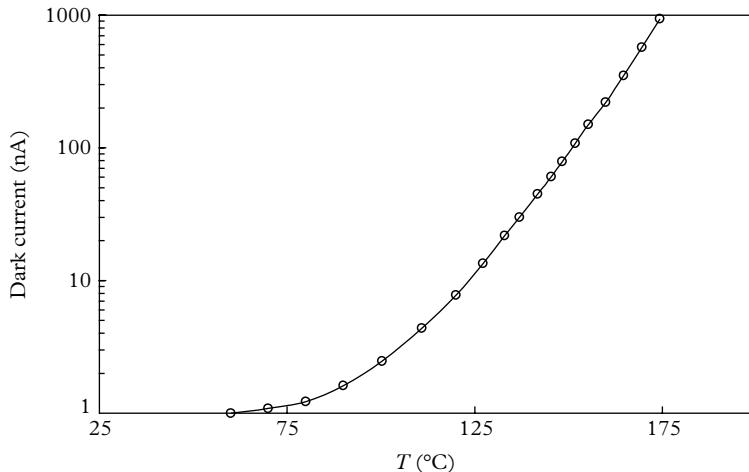


Fig. 12.9. The rapid rise in dark current with temperature is a major limiting factor affecting performance. Because of fatigue, the curve is not always repeatable after sustained operation at extreme temperatures.

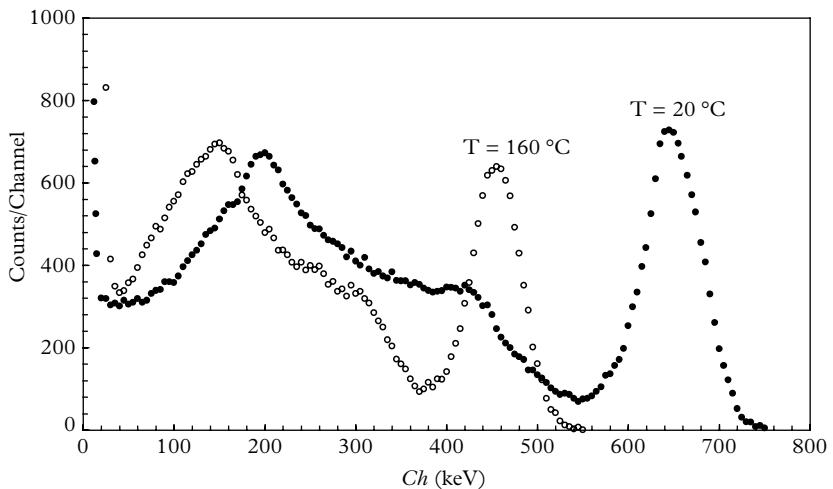


Fig. 12.10. Temperature induced reduction in sensitivity and resolution for ^{137}Cs and a sodium iodide detector. The energy scale on the abscissa refers to 20 °C.

edge, NE, appearing at low channel numbers. The NE provides a useful selection parameter; for example, it is 13 keV at 20 °C increasing to 26 keV at 160 °C. It is determined by the intercept of a horizontal line drawn from the Compton peak to the y-axis. Because of the steepness of the NE, the photopeak may also serve this purpose.

The accepted decay time constant of sodium iodide is 240 ± 10 ns at room temperature. However, the total light output produced by the scintillator degrades both for increases and for decreases in temperature, centred on room temperature. The decay schemes discussed by Ianakiev *et al.* (2009) reveal an unsound assumption of a single time constant for this scintillator: the decay becomes more rapid with increasing temperature (Rozsa *et al.* 1990). It is instructive to observe output pulses on an oscilloscope as they become narrower but without loss of pulse height as the temperature is increased. Oscilloscope traces in Fig. 12.11 show this reduction in the NaI(Tl) time constant from 240 to 100 ns at 160 °C. This can be turned to advantage, where a shaping amplifier is included in the detector, by obtaining performance similar to that in Fig. 12.12. The extension to the plateau translates into higher temperature operation by an additional ~ 10 °C. In the absence of a shaping amplifier, simply reducing the anode coupling capacitor produces a similar benefit. Note the additional advantage of handling faster count rates through a reduction in pulse overlap and dead time.

The role of an MCA in high-temperature applications is mainly for testing and development purposes, where exploration is based on gross counts only. An operating point is chosen at a suitably high PMT voltage to allow for the loss in overall sensitivity with temperature. This is illustrated in Fig. 12.13 for a pair of

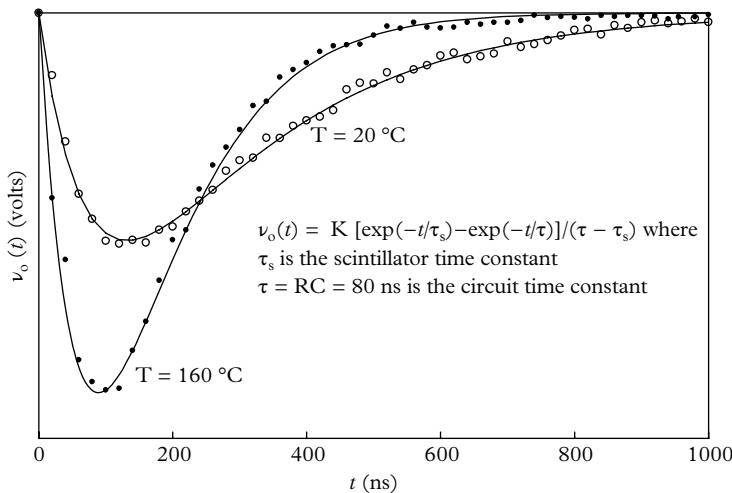


Fig. 12.11. Output signal pulse shapes from NaI(Tl) reveal a faster decay at high temperature, but not all users take advantage of this. The curves for $v_o(t)$ have the form of (14.4), with K a normalization constant to achieve equal areas.

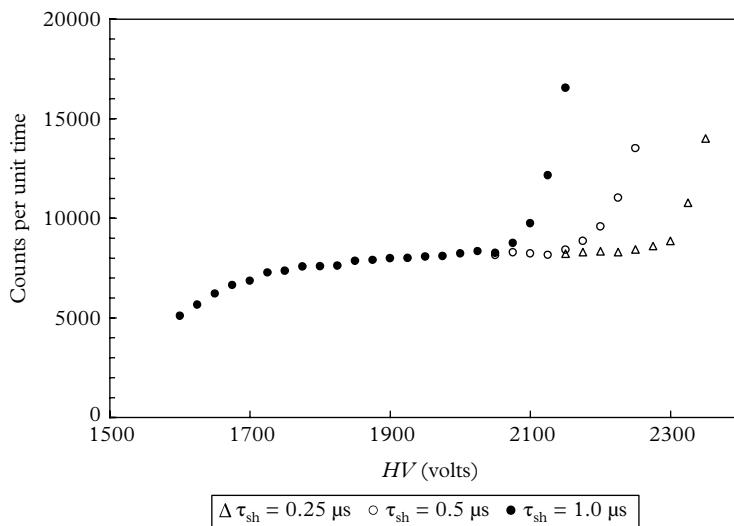


Fig. 12.12. The extension to the plateau characteristic for ^{137}Cs with three main amplifier shaping times, τ_{sh} , at 160°C .

plateau characteristics, with the box, in this instance, defined as the voltage span over which the count rate is constrained to within $\pm 2.5\%$ —in this case, 250 V. The choice of operating point should be selected anywhere between the centre of the box and the right-hand edge to mitigate detector fatigue. It is obvious that a

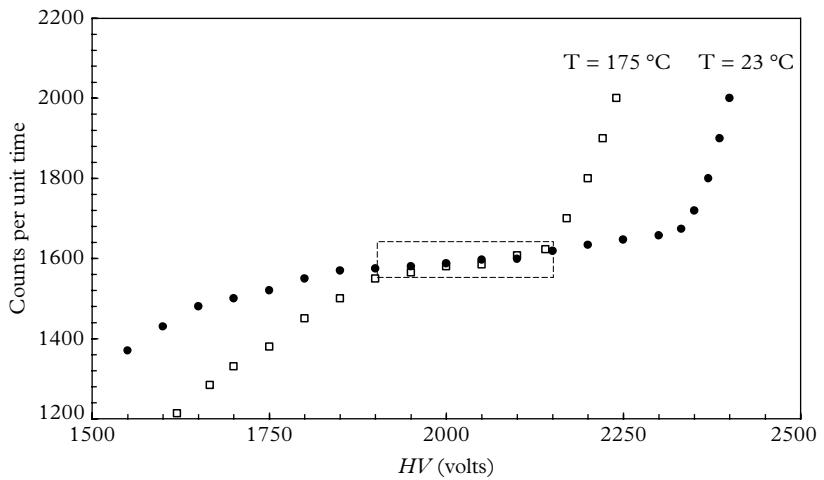


Fig. 12.13. Plateau characteristics for NaI(Tl) and a ^{137}Cs source. The breakaway from the plateau at high gain correlates with the noise edge.

plateau of low slope is desirable. The implication, since gain varies as V^n , is that a PMT with the minimum number of stages, n , may bring advantages.

REFERENCES

- Bristow, M. P., Bundy, D. H., and Wright, A. G. (1995). Signal linearity, gain stability, and gating in photomultipliers: Application to differential absorption lidars. *Appl. Opt.*, 34, 4437–52.
- Ianakiev, K. D. *et al.* (2009). Temperature behaviour of NaI(Tl) scintillation detectors. *Nucl. Instr. and Meth. in Phys. Res. A*, 607, 432–8.
- Krall, H. R. (1967). Extraneous light emission from photomultipliers. *IEEE Trans. NS*, 14, 455–9.
- Kume, S. *et al.* (1994). *Photomultiplier Tube*. Hamamatsu Photonics K K, Hamamatsu City.
- Lavoie, L. (1967). Photomultiplier cathode poisoning. *Rev. Sci. Instr.*, 38, 833–4.
- McAlpine, R. M. and Wright, A. G. (1994). Microphony in photomultipliers, *IEEE Conf. Rec.*, 1, 379–81.
- Philips. (1994). *Photomultiplier Tubes*, Pub. D PMT AB. Phillips Photonics, Brie, France.
- Persyk, D. E. *et al.* (1976). Photomultipliers for high-temperature applications. *IEEE Trans. NS*, 23, 185–8.
- Rozsa, C. *et al.* (1990). Characteristics of scintillators for well logging to 225 °C. *IEEE Trans. NS*, 37, 966–71.

13

Voltage dividers

13.1 Introduction

A PMT is said to be non-linear once the anode current ceases to be proportional to the intensity of the input light signal; non-linear performance can be reduced, but not entirely eliminated. Where voltage dividers are concerned, non-linearity refers to a relative change in gain, $\Delta g/g$, with anode current. The primary requirement of a voltage divider is to provide a set of biasing voltages on the dynodes and focussing elements of a PMT that remain fixed, regardless of changes in mean anode current. This is never quite achieved in practice with resistor-based dividers, as analysis will show. A voltage divider comprising a series of resistors is the preferred choice for most applications. Its appeal lies in its simplicity, low cost, and reliability. The default voltage divider consists of $(n + 1)$ identical resistors although improvements in performance can be gained by increasing the value of the $k-d_1$ resistor or replacing it with a zener diode. Where signals are pulsed, it is usual to increase back-end resistor values in a tapered manner: R , $1.5R$, $2R$, and $3R$.

Whichever variant of the default divider network is employed, amplification of the photocurrent interacts with the voltage-divider potentials causing a change in PMT gain. The log slope of the gain–voltage characteristic β influences gain instability, $\Delta g/g$. This is investigated in 13.2. The use of zener diodes to maintain fixed voltages, either at the front or back end of the divider, has a second-order effect of increasing β , although analysis in 13.5 will show that the overall improvement in $\Delta g/g$, over that achieved with a resistive divider, is still considerable. Uncovering the origin of non-linear behaviour in a voltage divider leads to the means for its alleviation. The word ‘unstable’, in the context of 13.3, refers to a voltage divider that fails to maintain constant interdynode voltages, and hence stable PMT gain, over a specified dynamic range of operation. Expressed simply, PMTs undergo a change in gain, originating in the divider, as the mean anode current is varied. Unless decoupling capacitors are included, a change in gain associated with transient currents of high peak value will also occur because of space charge effects. In addition, there is a form of non-linear behaviour intrinsic to PMTs, particularly those with BeCu dynodes, taking the form of a positive gain change with increasing anode current. These two intrinsic contributions are not included in the analysis followed in 13.3, but are treated separately. Voltage dividers designed for equipment powered by low-voltage rechargeable batteries (typically, 12 V) present a challenge: that of achieving acceptable linearity of performance with minimal power consumption.

The initial investigation concentrates on improving gain stability under variable anode current conditions. Analysis in 13.2 and 13.3 refers particularly to a uniform divider, comprising a common resistor value throughout. The performance for non-uniform dividers, in which front-end and back-end resistors may be different in value from the common resistor, can be deduced by adapting the analysis. Consideration is given in 13.6 to the inclusion of voltage

reference devices, such as zener diodes and voltage followers based on bipolar transistors, FETs or MOSFETs.

13.2 Gain–voltage relationships

13.2.1 Resistive dividers

The functional relationship between the multiplier gain, g , anode current, I_a , and HV, V , plays a key role in the performance of all voltage dividers. The overall gain is the product of n individual interdynode gains and, for analysis purposes, it is assumed that each stage is identical. The gain, δ , of a dynode is described by the semi-empirical relationship (13.1), in which a is a constant, v is the interdynode voltage, and α is an index less than unity:

$$\delta = av^\alpha. \quad (13.1)$$

As discussed in 5.2, $a = 0.36$, and $\alpha = 0.7$, are representative of a high-gain PMT. It is assumed in the analysis that these numerical values apply to all dynodes. Assuming otherwise adds needless complication, without leading to deeper understanding. When comparing actual performance against theoretical predictions, it is important not to lose sight of this basic assumption regarding (13.1)—all dynodes are assumed to behave identically. The intention of the analysis is to provide an insight regarding performance, rather than to account for exactly what is observed. It is customary in some applications to operate with higher voltages between the cathode and first dynode, and similarly at the back end, where interdynode voltages may be progressively increased—referred to as ‘tapering’. A parameter, m , is adopted for the initial analysis in Fig. 13.1 to characterize performance when $v(k - d_1)$ or $v(d_n - a)$ is varied. Many different divider distributions have been devised for specific PMT types and their applications, but the divider depicted in Fig. 13.1 is adopted for present purposes, leaving the reader free to adapt the analysis as required for other configurations.

From (13.1), the overall gain for an n -stage PMT, operated at fixed V , in the divider of Fig. 13.1(a) is

$$g = a(mv)^\alpha(av^\alpha)^{n-1},$$

where V and v are related as follows:

$$v = V/(m + n),$$

leading to

$$g = a^n m^\alpha [V/(m + n)]^{n\alpha}. \quad (13.2)$$

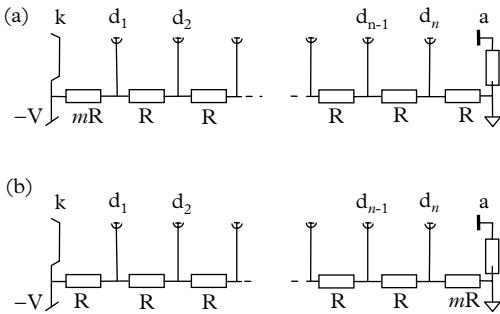


Fig. 13.1. The parameter, m , reveals the effect on gain of changing a resistor value; (a) $v(k - d_1)$ is varied and in (b) $v(d_n - a)$ is varied. Note, there is always one more resistor than there are dynodes. The distribution with $m = 1$ represents a default divider.

Differentiating (13.2) with respect to m reveals that $m = n/(n-1)$ generates maximum overall gain for a fixed HV. Examination of (13.2) reveals that the maximum gain condition applies wherever the variable resistor is placed, other than between d_n and the anode.

The gain for a uniform resistor network follows from (13.2) by setting m to unity:

$$g = a^n [V/(n+1)]^{n\alpha}. \quad (13.3)$$

The corresponding expression for gain, with a variable resistor located between the last dynode and the anode, as in Fig. 13.1(b), is

$$g = a^n \left(\frac{V}{m+n} \right)^{n\alpha}. \quad (13.4)$$

Equations (13.2) and (13.4) are illustrated in Fig. 13.2 for a ten-stage PMT, highlighting the important point that overall gain increases by reducing the last resistor in the divider. This is because there is no contribution to gain between d_n and the anode—the function of which is to collect charge generated by the last dynode. We verify in 13.3 that, since V is fixed, $v(d_n - a)$ decreases as signal is drawn from the anode, causing an increase in the remaining interdynode voltages, and hence an increase in overall gain. Appreciating this phenomenon is of fundamental importance and is further discussed in 13.3. Maximum gain, with a single, variable interdynode resistor, is attained for $m = n/(n-1)$, tending to unity as the number of stages is increased.

Interstage voltages $v(k - d_1), v(d_1 - d_2), v(d_2 - d_3), \dots, v(d_{n-1} - d_n)$ provide the energy for secondary electron generation, and $v(d_n - a)$ ensures collection of electrons from the last dynode. As shown, gain approaches a maximum, at fixed HV, when all interstage voltages, including $v(k - d_1)$, are identical. There are applications that call for increasing $v(d_n - a)$, in which case there is a decrease in overall gain. It will become evident that these preliminary observations and the formulations (13.1) to (13.4) are relevant to gain instability.

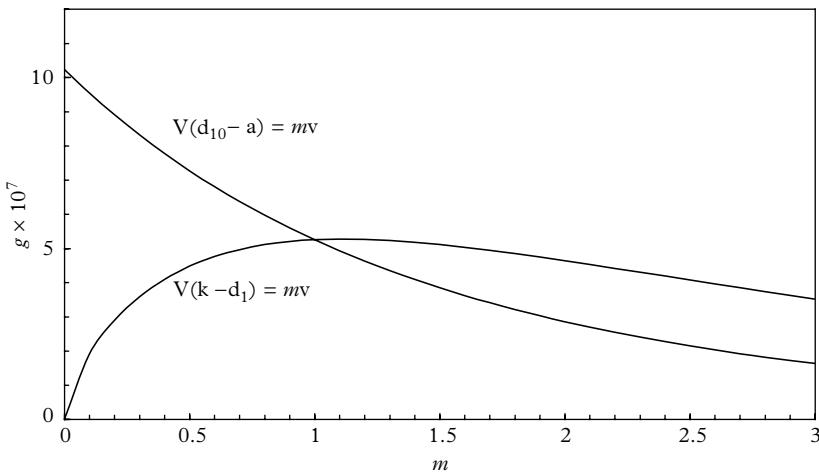


Fig. 13.2. Gain versus m for a high-gain ten-stage PMT characterized by $a = 0.36$, $\alpha = 0.7$ and operated at an HV of 600 volts. The curve marked $V(d_{10} - a)$ refers to (13.4) and that marked $V(k - d_1)$ refers to (13.2).

Consider the logarithmic form of (13.3). Since the gain–HV relationship is a power one, it follows that

$$\log g = \log \left(a^n / (n + 1)^{\alpha n} \right) + (\alpha n) \log V. \quad (13.5)$$

The first term on the right is a constant; hence,

$$\beta = \frac{\partial(\log g)}{\partial(\log V)} = \alpha n. \quad (13.6)$$

The log slope αn , referred to as β , ranges between 6 and 11, depending upon the type of secondary emission material and on the number of dynode stages, n . Note that all expressions for g so far discussed have the same log slope, β . Creating non-uniform dividers in accordance with (13.3) has the effect of shifting the gain–voltage curves along the abscissa, but always parallel to the curve representing a uniform divider.

13.2.2 Dividers with zener diodes

PMT performance is generally enhanced by operating with a higher voltage between the cathode and the first dynode. This applies particularly to low-gain applications concerning photoelectron collection, pulse height resolution, and speed of response. Improved performance can be achieved by simply increasing the value of R_{k-d_1} , or by substituting it with a zener diode. The latter provides the additional benefit of a fixed voltage, independent of HV, and hence stable collection efficiency. Zener diodes are also recommended for certain applications,

in place of the resistor between the last dynode and the anode. The reason for this is discussed in 13.3.

Replacing R_{k-d_1} by a zener diode leads to

$$g = av_z^\alpha (av^\alpha)^{n-1} = a^n v_z^\alpha \left(\frac{V - v_z}{n} \right)^{\alpha(n-1)}, \quad (13.7)$$

$$\beta = \frac{\partial(\log g)}{\partial(\log V)} = (n-1)\alpha V / (V - v_z). \quad (13.8)$$

If the resistor between the last dynode and the anode is replaced by a zener diode, then

$$g = a^n \left(\frac{V - v_z}{n} \right)^{\alpha n}, \quad (13.9)$$

$$\beta = \frac{\partial(\log g)}{\partial(\log V)} = n\alpha \left(\frac{V}{V - v_z} \right). \quad (13.10)$$

In (13.8) and (13.10), β is a function of V , showing the effect on the gain of including zener diodes. Small-value zener diodes (< 100 V) have little influence, as is clear by comparing curves (a) and (b) in Fig. 13.3. But it is obvious that the inclusion of a series of zener diodes, totalling, say, 300 V, will cause a significant reduction in gain, particularly at low to medium V . This can be restored simply by increasing V but the enhanced slope, clearly shown in Figs 13.3 and 13.4, implies an increased

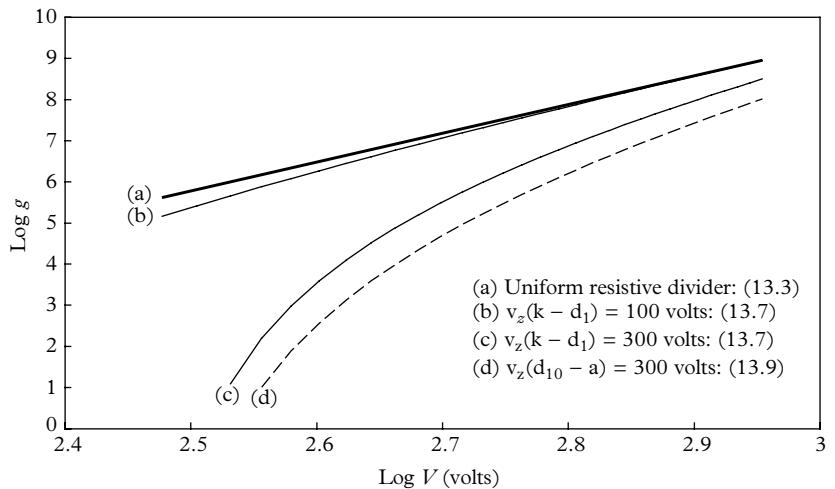


Fig. 13.3. Gain–voltage relationships for various configurations: (a) a uniform divider comprising resistors only (see (13.3)); (b) a 100 V zener diode replaces the first resistor of the divider (see (13.7)); (c) as for (b) but $v_z = 300$ volts (see (13.7)); (d) $v_z = 300$ volts, located between d_{10} and the anode (see (13.9)).

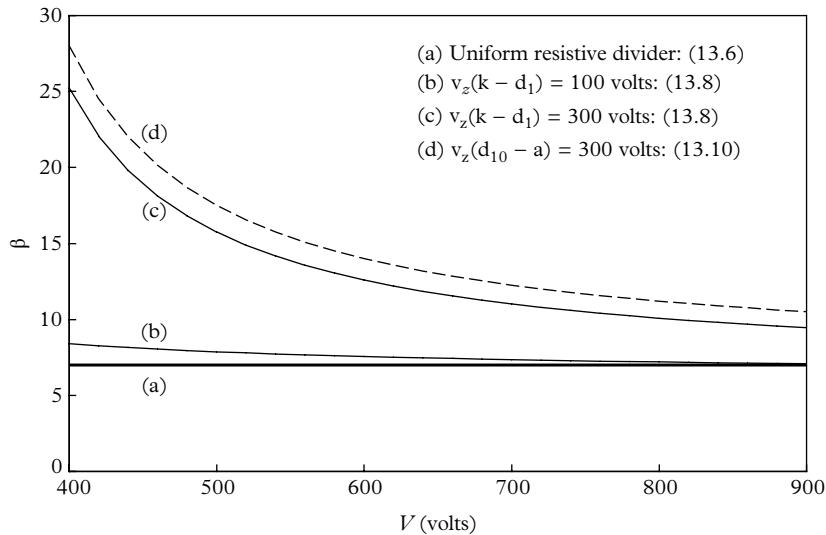


Fig. 13.4. The log slope of the gain–voltage characteristic for the PMT depicted in Fig. 13.3. Curve labelling follows that in Fig. 13.3.

dependence upon the stability of the HV supply, V . Quality power supplies should meet this requirement, however, zener diodes may contribute to long term instability because v_z is temperature sensitive.

13.3 Analysis of resistive dividers

Formulations (13.2) to (13.10) are derived in the absence of signal current, that is, I_k and I_a are zero. Investigations into linearity of response require equations relating these two currents to the gain, g . The PMT is operating linearly when the ratio I_a/I_k is constant as I_a is varied. The analysis in this section is subject to the following assumptions: there are n stages of gain and, consequently, $(n + 1)$ interdynode voltages; all resistors have the same value, R ; dynode secondary emission coefficients, δ , are identical for all dynodes. It will be obvious from the complexity of the relationships developed on this basis that any generalization of these parameters will incur unwieldy formulations without providing further insight. We assume fixed illumination, resulting in a cathode current I_k and a corresponding anode current, I_a . These two parameters are related through $I_k = I_a/\delta^n$, where δ^n is the gain of the multiplier, g . Pulsed-light applications are covered, but only with respect to the mean current so generated. Pulse height linearity on a pulse to pulse basis is treated in 13.7. Kirchhoff's law leads to the current division shown in Fig. 13.5.

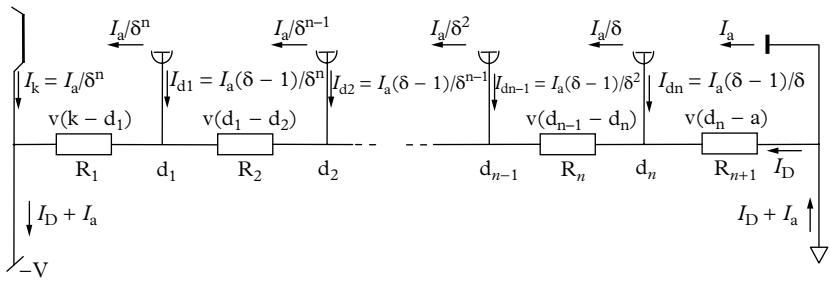


Fig. 13.5. A resistive divider with common resistor values; R_1 is the first resistor of the string, and R_{n+1} is the ultimate one. Note that currents of the form $I_a/\delta^n, \dots, I_a/\delta^{n-1}$ are internal to the PMT and cannot be measured directly. The current drawn by the supply, $I_D + I_a$, is mirrored in the ground leg.

It will become clear that I_D , defined as the current flowing into R_{n+1} , is a key parameter in understanding the behaviour of a voltage divider and, in particular, its relationship to I_a ; I_D flows through all the resistors in the string and it is progressively supplemented by dynode currents $I_a(\delta - 1)/\delta, I_a(\delta - 1)/\delta^2, \dots, I_a(\delta - 1)/\delta^n$ as each dynode is passed. The analysis is based on a snapshot of the contributions to steady-state current flow, in order to derive an expression for the current through R_1 . By inspection, it is obvious that all dynode currents flow through R_1 , but not the cathode current, I_a/δ^n . Hence, the current I_{R1} through R_1 , except the cathode current, is $I_D + I_a - I_a/\delta^n$. Similarly, I_{R2} comprises $I_D + I_a - I_a/\delta^n - I_a(\delta - 1)/\delta^{n-1}$ since in this case I_k and I_{d1} do not contribute. Continuing this process leads to expressions for the following resistor currents:

$$\begin{aligned} I_{R1} &= I_D + I_a - I_a/\delta^n, \\ I_{R2} &= I_D + I_a - I_a/\delta^{n-1}, \\ I_{R3} &= I_D + I_a - I_a/\delta^{n-2}, \\ &\dots \\ I_{R(n-1)} &= I_D + I_a - I_a/\delta^2, \\ I_{Rn} &= I_D + I_a - I_a/\delta, \\ I_{R(n+1)} &= I_D. \end{aligned} \quad (13.11)$$

Interstage voltages are simply R times these currents, assuming common resistor values. In the absence of anode current, all interstage voltages are common and equal to $V/(n + 1)$. Once anode current is drawn, resistor currents change in accordance with (13.11). Interstage voltages at the back end decrease while those nearer the photocathode increase, as shown in Figs 13.6 and 13.7, and the net effect is an increase in overall gain with anode current.

Interstage voltages sum to V , and we have, from (13.11),

$$\begin{aligned} \frac{V}{R} &= \left[(n + 1)I_D + nI_a - I_a \left(\frac{1}{\delta} + \frac{1}{\delta^2} + \dots + \frac{1}{\delta^n} \right) \right], \\ \frac{V}{R} &= (n + 1)I_D + nI_a - \frac{I_a}{\delta} \left(1 + \frac{1}{\delta} + \frac{1}{\delta^2} + \dots + \frac{1}{\delta^{n-1}} \right). \end{aligned} \quad (13.12)$$

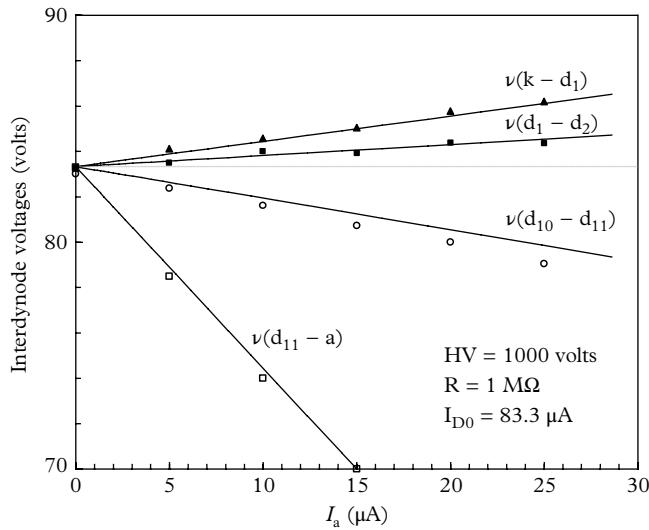


Fig. 13.6. Drawing anode current causes rearrangement of interstage voltages in an 11-stage PMT. The divider is uniform, consisting of $1 \text{ M}\Omega$ resistors, and the predictions are based on (13.16). Curves for $v(d_2 - d_3)$ to $v(d_9 - d_{10})$ fall between those for $v(d_1 - d_2)$ and $v(d_{10} - d_{11})$. Measured stage voltages with a multimeter are shown by the points.

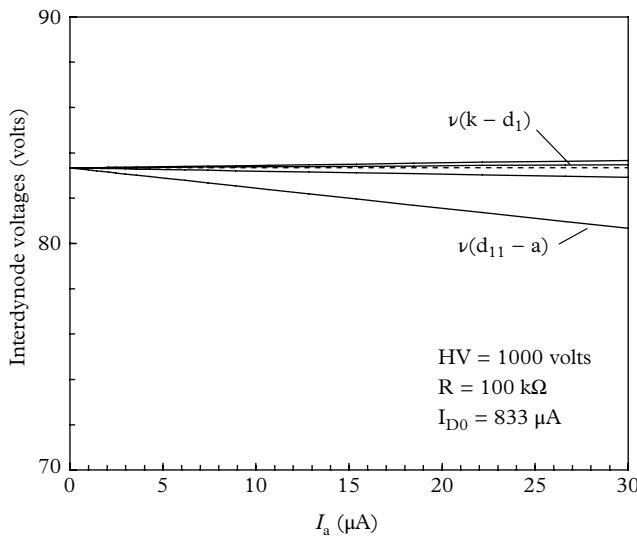


Fig. 13.7. The result of using lower resistor values of $100 \text{ k}\Omega$ is a reduced dependence on anode current. The sequence of curves is the same as in Fig. 13.6.

The geometric series sums to

$$\text{sum} = \frac{\delta^n - 1}{\delta^n} \cdot \frac{\delta}{\delta - 1}$$

and note that $\delta^n \gg 1$, $[(\delta^n - 1)/\delta^n]$ is effectively unity and hence (13.12) becomes

$$\begin{aligned} V/R &= (n+1)I_D + nI_a - I_a(\delta - 1) \\ &= (n+1)I_D + (n+1)I_a - I_a - \frac{I_a}{\delta - 1} \\ &= (n+1)I_D + (n+1)I_a - I_a \left(1 - \frac{1}{\delta - 1}\right) \\ &= (n+1)I_D + (n+1)I_a - I_a \frac{\delta}{\delta - 1}; \\ \text{therefore } \frac{V}{(n+1)R} &= I_D + I_a - \frac{\left(\frac{I_a \delta}{\delta - 1}\right)}{n+1}. \end{aligned} \quad (13.13)$$

Thus, $V/[(n+1)R]$ is the bias current through the divider in the absence of anode current, referred to as I_{D0} , and we have

$$I_D = I_{D0} - I_a + I_a \frac{\left(\frac{\delta}{\delta - 1}\right)}{(n+1)}, \quad (13.14)$$

and

$$I_{HV} = I_D + I_a = I_{D0} + I_a \frac{\left(\frac{\delta}{\delta - 1}\right)}{(n+1)}. \quad (13.15)$$

Substituting (13.14) into (13.11), we obtain

$$\begin{aligned} I_{R1} &= I_{D0} + \frac{I_a \delta}{(\delta - 1)(n+1)} - I_a / \delta^n, \\ I_{R2} &= I_{D0} + \frac{I_a \delta}{(\delta - 1)(n+1)} - I_a / \delta^{n-1}, \\ I_{R3} &= I_{D0} + \frac{I_a \delta}{(\delta - 1)(n+1)} - I_a / \delta^{n-2}, \\ &\dots \\ I_{R(n-1)} &= I_{D0} + \frac{I_a \delta}{(\delta - 1)(n+1)} - I_a / \delta^2, \\ I_{Rn} &= I_{D0} + \frac{I_a \delta}{(\delta - 1)(n+1)} - I_a / \delta, \\ I_{R(n+1)} &= I_{D0} + \frac{I_a \delta}{(\delta - 1)(n+1)} - I_a. \end{aligned} \quad (13.16)$$

Since the currents in the resistors change, so will the interdynode voltages. The change in $I_{R(n+1)}$, with respect to I_a , is in practice always negative since $\delta/[(\delta - 1)(n + 1)]$ is less than unity. Typically, for $\delta = 4$, and $n = 11$, $\delta/[(\delta - 1)(n + 1)] = 1/9$. It is clear that I_{Rn} also decreases for the assumed parameters, but $I_{R(n-1)}$ and the remaining resistor currents all increase. The set of equations (13.16) predicts the change in interdynode voltages with anode current and are comparable with measurements in Figs 13.6 and 13.7 for the stated parameters n , I_{D0} , and δ . Furthermore, (13.14) and (13.15) indicate that

$$I_D \approx I_{D0} - 0.9 \times I_a, \\ I_{HV} \approx I_{D0} + I_a/10.$$

The current drawn from the supply increases from the quiescent setting in proportion to a small fraction of I_a and, from (13.14),

$$\Delta g/g = \frac{(I_{D0} - I_D)}{I_{D0}} = \frac{I_a}{I_{D0}} \left[1 - \frac{\delta}{(n + 1)(\delta - 1)} \right]. \quad (13.17)$$

The change in gain varies approximately as I_a/I_{D0} and to maintain stable gain within 1% therefore requires a standing divider current 100 times the anode current; this is the basis for the long-established adage for PMT biasing. Equation (13.17) is also given in the Philips (1980) handbook, but without derivation, and by Land (1971) using a different approach. Increasing I_{D0} extends the quality of linear performance proportionately but there is a limit of about 1 mA on the maximum value of the standing current, if heating effects are to be avoided.

Taking typical values of $n = 11$, and $\delta = 4$, (13.14) and (13.15) may be further approximated to

$$I_D \approx I_{D0} - I_a, \\ I_{HV} \approx I_{D0}, \quad (13.18)$$

respectively. The current drawn from the supply is approximately constant, where anode currents are modest—an important result. The voltage drop in $\nu(d_n - a)$ is given by

$$\nu(d_n - a) \approx R(I_{D0} - I_a), \quad (13.19)$$

indicating that $\nu(d_n - a)$ decreases approximately linearly with increasing I_a . Returning to (13.3), we obtain

$$g = \frac{a^n V^{an}}{(n + 1)^{an}} = bV^\beta, \quad (13.20)$$

where $\beta = \alpha n$ by definition, and b is a constant. The relative gain change from (13.20), using (13.19) is

$$\frac{\Delta g}{g} = \frac{\beta \Delta V}{V} \\ = \frac{\beta R I_a}{V}. \quad (13.21)$$

This is essentially the same expression as that given by Lush (1965) and Piétri (1981). Since the last resistor does not contribute to gain, and V is fixed, a drop in $\nu(d_n - a)$, as a result of drawing anode current, is redistributed across the remaining dynodes; this is illustrated in Figs 13.6 and 13.7 for an 11-stage linear focussed ET 9107B. Drawing anode current has the effect of increasing the HV by ΔV , with an accompanying increase in gain, given by (13.21); $\Delta g/g$ is always underestimated by (13.21) because $\nu(d_{n-1} - d_n)$ also decreases with anode current (see Fig. 13.6). This is a second-order effect, and the contribution to gain change from a decrease in $\nu(d_{n-1} - d_n)$ has not been included in arriving at (13.21). This equation can be expressed in a different form by substituting R with $V/[(n + 1)I_{D0}]$ and assuming n is sufficiently large that it may be taken as $(n + 1)$; thus,

$$\frac{\Delta g}{g} = \frac{I_a}{I_{D0}} \frac{\beta}{(n + 1)}, \quad (13.22)$$

which is identical to (13.21), but expressed in different parameters. The two entities, β and n , are roughly of the same order of magnitude and, as before, the relative change in gain varies approximately as I_a/I_{D0} . The expression in (13.22) is also quoted by Piétri (1981) and Flyckt and Marmonier (2002). An important point to note is the linear dependency of $\Delta g/g$ on β .

13.4 Resistive divider operation

There is an alternative method for determining $\Delta g/g$, as a function of I_a , by substituting the interstage voltages based on (13.1) to obtain (13.23):

$$g(I_a) = a[\nu(k - d_1)]^\alpha \times a[\nu(d_1 - d_2)]^\alpha \times a[\nu(d_2 - d_3)]^\alpha \times \cdots \times a[\nu(d_{n-1} - d_n)]^\alpha$$

therefore $\Delta g/g = \frac{g(I_a) - g(0)}{g(0)}$ (13.23)

where $g(0) = bV^\beta$.

Interdynode voltages follow from (13.16) for substitution into (13.23). The integral non-linearity can be deduced by using closely spaced data. An example, following this procedure, is given in Fig. 13.8 together with the prediction of (13.21). We note the prediction of (13.21) is linear with I_a but is an underestimate compared to that from (13.23). The predictions of both formulae are acceptable, given the assumptions. It is clear from (13.21), and from direct measurement, that linearity with a resistive divider may be approached but is never achieved, as previously stated. Satisfactory performance, however, can be achieved in terms of establishing an upper limit to non-linearity. For example, substituting the parameters and settings quoted in Fig. 13.8 into (13.21) predicts that integral non-linearity of less than 0.7% can be achieved for anode currents smaller than 1

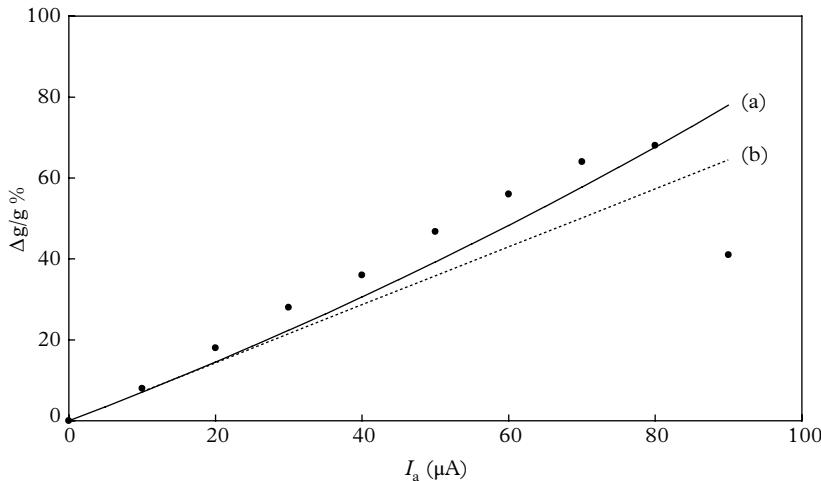


Fig. 13.8. Gain change as a function of anode current for a uniform divider with a quiescent current of $100 \mu\text{A}$; $R_1 - R_{12} = 620 \text{ k}\Omega$; $\beta = 8.64$; and $V = 750 \text{ V}$: (a) using the method based on (13.23); (b) using the method based on (13.21). The agreement with experimental points, considering the assumptions, is surprisingly good. The last two experimental points indicate folding as I_a approaches I_{D0} . Because of the assumption made in (13.19), (b) predicts a smaller change in relative gain than (a).

μA . This figure can be extended to $10 \mu\text{A}$, for example, by increasing the bias current tenfold. Alternatively, or in addition to this, it is always possible to correct readings, as discussed in Chapter 9.

Equation (13.21) predicts that the magnitude of the gain DC gain change, $\Delta g/g$, for a resistive divider varies inversely with V but linearly with β , I_a , and R_{n+1} . Non-linearity with a uniform divider is shown in Fig. 13.9, together with the performance of a tapered divider, generally used in detectors for high-energy physics. The latter is designed to extend the dynamic range in pulsed applications, but at the expense of serious gain shift at high count rates. Dependence on the magnitude of the last resistor is easily verified by adopting a modified divider with $\nu(d_n - a)$ reduced. For the configuration in Fig. 13.9(c), $R_{n+1} = R/3$, with $\Delta g/g$ decreasing by a factor of about 3 for all I_a , as predicted. For anode currents exceeding $50 \mu\text{A}$, there is a flattening followed by a dip in the curve that is possibly a result of space charge effects, exacerbated by the relatively low collection voltage applied between the anode and the last dynode.

Applications involving pulsed-light sources often demand a wide dynamic range. This implies the ability to handle both small and large pulses with the same PMT. Small pulses cause no difficulty, whereas signals at the upper end of the dynamic range may suffer from the effects of space charge saturation (see 5.6.5). The pulse height at which non-linear gain is noticeable in a resistive divider may be extended by increasing the back-end voltages by using a tapered divider.

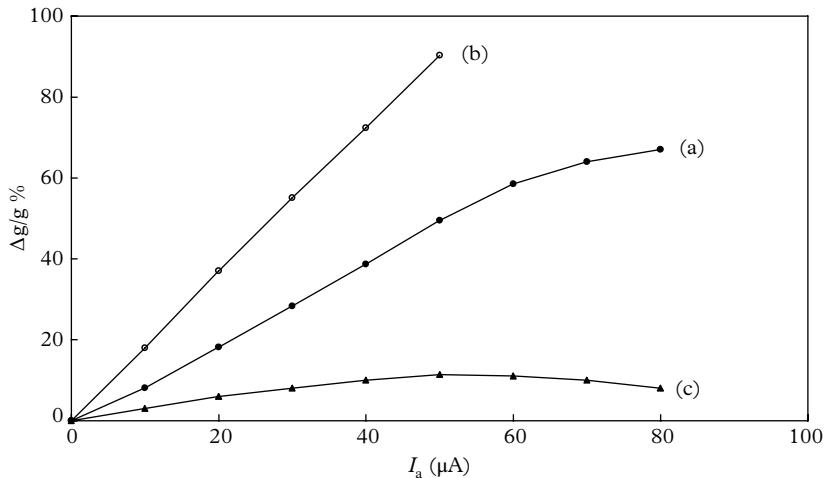


Fig. 13.9. Measured linearity for an 11-stage 9107B, a high-gain PMT, in different dividers: (a) uniform divider: R, R, \dots, R, R ; (b) tapered divider: $R, R, \dots, R, 2R, 3R, 4R, 3R$; (c) Anti-taper: $R, R, \dots, R, R/3$. Data refers to $R = 620 \text{ k}\Omega$, and $I_{D0} = 100 \mu\text{A}$.

The degree of acceptable taper depends on the particular electron-optical design for the PMT in question but the arrangement presented in Fig. 13.9, terminating in $3R$, is typical of designs used in high-energy physics. Manufacturers provide recommended divider configurations for specific pulsed applications. Supre-linearity generated by tapered dividers is sometimes tuned to partially cancel negative pulse height non-linearity, invariably providing a means to extend the dynamic range to meet a difficult specification—this is not recommended where signal amplitudes and count rates are variable. Physicists refer to the cause of gain shift as a rate effect because of the observed link between count rate and gain change. Even today, it is not generally realized that gain change is primarily related to mean anode current and to the voltage divider. In pulsed applications, for example in scintillation spectroscopy and in photon counting, the mean anode current is clearly proportional to the rate of events, so the two interpretations are compatible.

13.5 Inclusion of zener diodes

It was shown in the previous section that the major source of gain instability in resistive dividers stems primarily from the variation in the voltage between the last dynode and the anode—a consequence of generating anode current. It is obvious that improved performance in an n -stage PMT should follow by maintaining

$v(d_n - a)$ constant, and thus independent of changes in anode current. Replacing R_{n+1} in Fig. 13.5 with a zener diode is one way of achieving this. Although the overall voltage between the cathode and the last dynode is now fixed, there is a partial redistribution of $v(d_{n-1} - d_n)$ among the earlier dynodes in the manner of (13.16), once the anode current is drawn. Circuit analysis similar to that presented in 13.4 may be pursued but it is of limited use. There are two considerations that cast doubt concerning this particular variant: zener diodes have finite slope resistance, which should be included in any modelling. Second, the gain of individual dynodes varies from one dynode to the next in an unknown manner. The author's experience is that performance is improved by the use of a zener diode but that $\Delta g/g$, although reduced, may be positive or negative depending on the number of zener diodes involved. Measurements taken with zener-stabilized back-end voltages are presented in Fig. 13.10, showing considerable improvement over the performance of a uniform, resistive divider.

It is customary to operate a PMT with $v(k - d_1)$ set higher than the interstage voltages—typically, by a factor of 3. This improves single and multi-photoelectron resolution, collection efficiency, and transit-time dispersion. Alternatively, one or more zener diodes, connected in series, may be used in place of R_1 in Fig. 13.5, giving independence from the overall HV setting. According to (13.4), one or more resistor values may be changed with little effect on β or gain. However, exchanging a resistor for a zener diode has a different outcome: in this case, β changes with V , especially noticeable at low V . This is described by (13.8) and

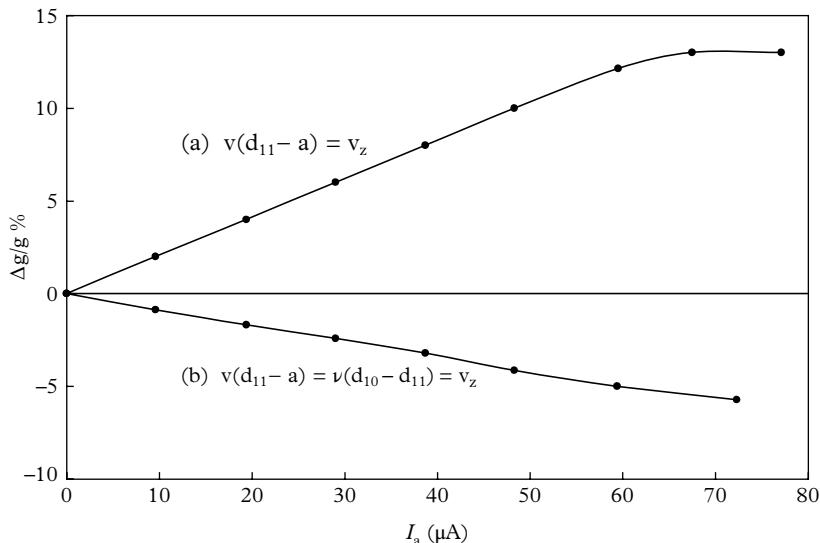


Fig. 13.10. Performance of an 11-stage, 9107B, PMT with zener-stabilized stages. $R = 620 \text{ k}\Omega$; $I_{D0} = 100 \mu\text{A}$; and $v_z = 60$ volts. These curves should be compared with those of Fig. 13.9 while noting the different scale on the ordinate.

(13.10) and illustrated in Fig. 13.4. According to (13.21), the penalty for operating with high β is enhanced gain shift with anode current. This is a particular concern with large-diameter PMTs, such as a fast 5" and hemispherical PMTs, where the dividers invariably include zener diodes. It is usual with these PMT types to use a series of zener diodes, totalling up to 1000 V, to achieve the desired $v(k - d_1)$; the consequential sensitivity to changes in V can be gauged from Figs 13.3 and 13.4. Equally important is the enhanced gain sensitivity of $\Delta g/g$.

Improvement in gain stability can be achieved by including a single zener diode at the back end of a divider, but the choice of v_z must be compatible with neighbouring interstage voltages. This is particularly critical with linear focussed dynodes because collection at the anode is affected by any irregularity in the neighbouring field patterns; I_D in this case is approximately $I_{D0} - I_a$ and it must be chosen greater than the expected maximum anode current, $I_a(\max)$, to avoid cut-off. In fact, I_{D0} should be greater than $I_a(\max) + I_z(\text{plateau})$, where the latter is the minimum current over which zener diodes are able to provide stable voltage—about 25 μA for the device depicted in Fig. 13.11. Note that this requirement does not apply to a zener diode located at the front end, because the zener current is, according to (13.16), always greater than I_{D0} , ensuring against cut-off. An ideal zener diode, for voltage-divider purposes, is one with the following properties: (i) lowest possible turn-on current, to suit highly resistive voltage dividers, and (ii) low noise. Zener diodes are avalanche devices and hence they are inherently noisy, as shown in Fig. 13.12. This wideband noise may be observed at the output, particularly if the zener diode is located in proximity to the anode. Pick-up from zener diodes used to stabilize $v(k - d_1)$ is less likely, but it sometimes occurs. The solution to this problem lies in reducing the noise by the addition of passive

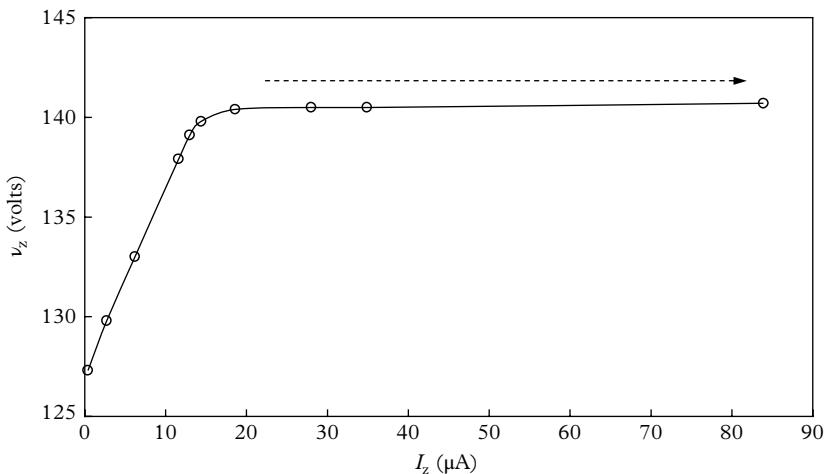


Fig. 13.11. Zener diode current–voltage characteristic for a BZT03C150. The arrow indicates the region of acceptable performance.

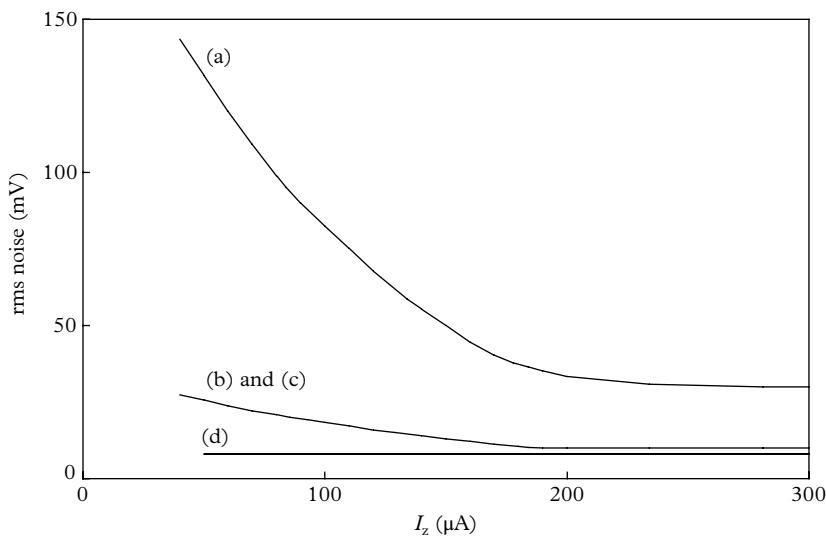


Fig. 13.12. Noise, measured with a 200 MHz bandwidth oscilloscope as a function of current, I_z , in a BZT03C150, 150 V, zener diode. The labelling of the curves follows that in Fig. 13.13. Noise refers to the terminals of the circuits when connected to a load resistor of $1 \text{ M}\Omega$.

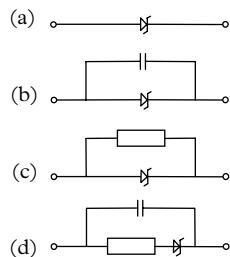


Fig. 13.13. Means for reducing noise from zener diodes. Configurations (b) or (c) may prove adequate for reducing noise, with $C = 10 \text{ nF}$, and $R = 1 \text{ M}\Omega$, whereas (d) provides optimal noise reduction, with typically $10 \text{ k}\Omega$, and 10 nF .

components, as in Fig. 13.13. Configuration (d) is most effective, but smoothing is achieved by compromising its voltage stabilization fidelity through an increase in slope impedance. For battery-operated equipment, it is important to select a device that requires minimal current to stabilize—a BZT03C150, for example, demands at least $20 \mu\text{A}$, implying that it would be suitable for use in a divider with common resistor values of up to $5 \text{ M}\Omega$, and operated at around 1000 V overall. The use of high-impedance resistors, or low HV, implies minimal I_{D0} and, consequently, limited gain stability.

Zener diodes are temperature sensitive: devices with $v_z < 5.6$ V have a negative temperature coefficient, while those of higher v_z show an increase in stabilization voltage with temperature; this can be as much as 0.5 to 1.0 V/ $^{\circ}$ C for a 200 V device. The temperature stability of a divider with a zener diode at the front end will generally be acceptable, since a change in v_z voltage is redistributed among the other dynode potentials. Compensation, using a combination of low-voltage and HV zener diodes is possible but seldom employed. There is considerable dispersion in v_z , even at constant temperature, and among devices of the same type: it is not unusual to find $\pm \Delta v_z$ of the order of 5 to 10 V in any unselected batch of 200 V devices, for example.

13.6 Active voltage dividers

Circuits published in the 1970s by Kerns (1977), Hiebert *et al.* (1977), and Takeuchi and Nagai (1985), in articles entitled ‘A high-rate phototube base’, ‘Photomultiplier tube base for high pulsed anode currents’, and ‘Low power photomultiplier base circuit’ respectively, point to the problem already highlighted: that is, how to maintain fixed voltages on the dynodes of a PMT in the presence of a variable signal rate. Performance improvement with zener diodes has, apart from those already mentioned, the disadvantage that stabilized inter-stage voltages do not scale with HV. Another innovation, known as a ‘booster’, employs two HV supplies: one serving the back end of a divider, and the other at the front end. This mode of operation, developed by Duteil *et al.* (1977), recognizes that a high-standing current is only necessary over the last three or four dynodes, and subdividing the HV in this way leads to reduced power consumption. Despite air ventilation, these boosters ‘run hot’ and are not entirely satisfactory. Improving on the performance calls for a divider configuration with the following properties:

- low power consumption
- biasing voltages that track the HV
- biasing voltages that are constant with respect to variable anode current

Such performance can be realized with voltage dividers consisting of either NPN or PNP bipolar transistors or, equivalently, with N-type or P-type FETs and MOSFETS. Symbolic representation of FETs can be confusing, and an equivalent representation, in terms of the more familiar bipolar transistor symbols, is given in Fig. 13.14. The arrow within the transistor symbol indicates the direction of current flow, knowledge of which always helps when analysing circuits.

A fully active divider consists of two parallel circuits: a resistive one defines the required voltage distribution, and the other consists of a series of high-impedance

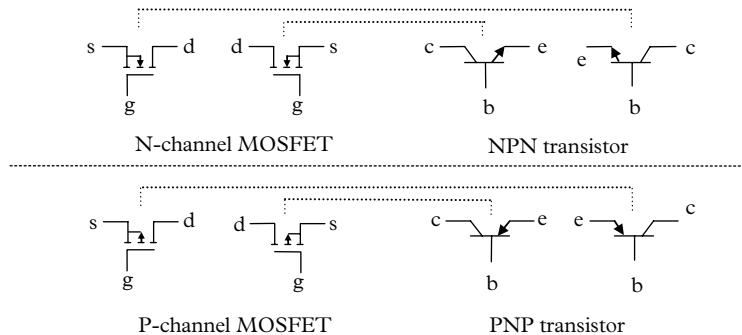


Fig. 13.14. Equivalent representations of active devices, indicated by the dotted lines. The arrow shows the direction of current flow in bipolar transistors.

active devices arranged to pick off these voltages. It is obvious for the circuit described in Fig. 13.15 that the potentials on a series of divider resistors, connected in parallel with active devices, will not change with anode current, provided that the active devices have sufficiently high input impedance. This prerequisite is easily met by MOSFETs and to a restricted extent by bipolar transistors. The use of active components thus leads to the attainment of an ideal voltage divider. The base-emitter voltage of a properly biased junction transistor is of the order of 0.7 V, essentially fixed, and independent of collector current. The same considerations apply to FETs but the gate-source voltage in these devices approaches a few volts. The key property of a voltage follower is the following: if the base voltage is constant, so also is the emitter, and similarly for FETs. Consider, as in Fig. 13.15, an NPN or N-type divider, in which I_{D0} is the current flowing through the transistor string in the absence of signal. In the analysis that follows, we assume perfect voltage followers with infinite input impedance and, in addition, V_{be} and V_{gs} are ignored. This isolation of the resistive string by an active one means that only the resistor current sets the stage voltages, without participation in the dynamics of the PMT gain.

13.6.1 Fully active N-type analysis

The analysis that follows, based on Fig. 13.15, assumes ideal active devices configured as voltage followers. No current is drawn from the resistor string to supply I_a and the voltage across R_B is fixed and equal to $V/(n + 1)$. Since this voltage is fixed, so is I_{D0} , the current through R_B ; I_D plus dynode currents $I_a(\delta - 1)/\delta$, $I_a(\delta - 1)/\delta^2$, ..., $I_a(\delta - 1)/\delta^n$ sum to $I_D + I_a - I_k$ through R_B . The current flowing into the HV supply is $I_D + I_a + I_R$ and, since $I_{D0} = I_D + I_a - I_k$,

$$I_{HV} = I_{D0} + I_k + I_R. \quad (13.24)$$

In practice, the photocurrent, I_k , is always negligible compared with the bias current: consequently, I_{HV} is also unaffected by a variable anode current, since it

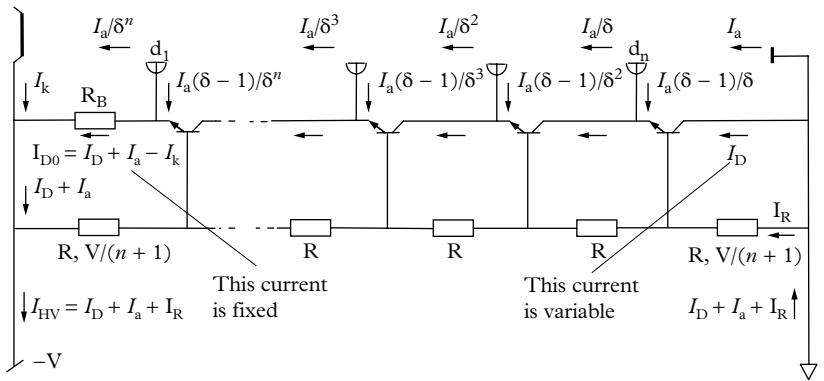


Fig. 13.15. Showing the current division in an N-channel, fully active divider. The voltages on the dynodes are fixed because of the voltage follower action of the active devices; I_R is the current through the resistor string. Note $I_D = I_{D0} - I_a + I_k$, and it follows that $I_{HV} = I_{D0} + I_k + I_R$.

does not appear in (13.24). Dynamically, as is obvious in Fig. 13.15, I_D changes in antiphase with I_a . Resistors marked R are normally of high impedance, since they are buffered from the PMT by active devices and thus able to maintain a fixed voltage—10 MΩ is quite common.

Understanding the action of this divider is more readily achieved by analysing the performance of a two-stage PMT, without loss of generality, avoiding the needless complication introduced by assuming an n-stage divider. The flow of currents corresponding to drawing a 1 μA and a 100 μA anode current is shown in Fig. 13.16.

13.6.2 Fully active P-type analysis

The analysis of a P-channel divider in Fig. 13.17 is similar to that for an N-channel divider, except the bias resistor is located at the anode:

$$I_{HV} = I_{D0} + I_a + I_R. \quad (13.25)$$

Here, I_{HV} varies linearly with respect to I_a , demonstrating that supply current is drawn on demand. The two constant currents, I_{D0} , and I_R , may be set as low as 10 μA each when using MOSFETs, making this divider ideally suited to battery-operated equipment. The current through R_B is the bias current, I_{D0} , given by $[V/(n+1)]/R_B$. Figure 13.18 facilitates understanding by analysing the behaviour of a two-stage P-type circuit for anode currents of 1 μA and 100 μA.

Equations (13.24) and (13.25) are illustrated in Fig. 13.19, together with I_{HV} for a resistive divider.

The current drawn (in microamps) from the three power supply types, for the conditions specified in Figs 13.15 and 13.17, is as follows:

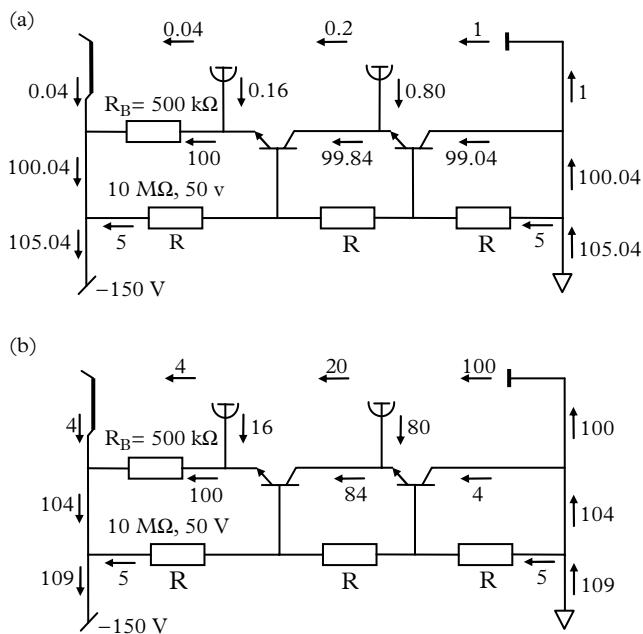


Fig. 13.16. Illustrating the current division in an N-type divider with assumed stage gains of $\delta = 5$. Note that the circuit limits at anode currents approaching $100 \mu\text{A}$ in this example. All currents are in microamps. Performance is shown for an anode current of $1 \mu\text{A}$ in (a), and $100 \mu\text{A}$ in (b). Note that the circuit ceases to function for $I_a > 104 \mu\text{A}$; extending the analysis to n stages is obvious.

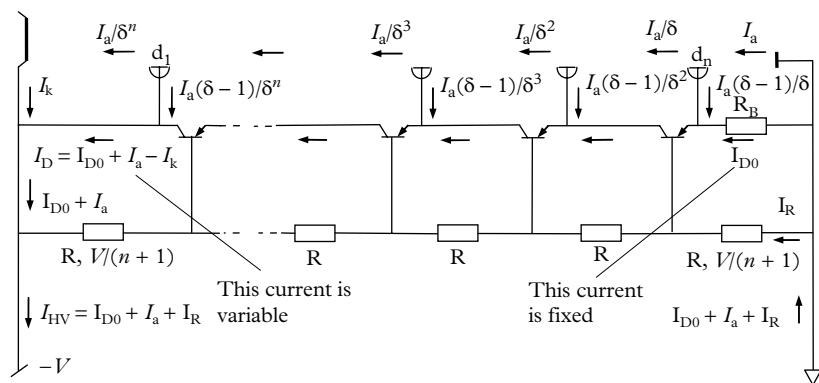


Fig. 13.17. Showing the current division in a P-channel, fully active divider.

Resistive from (13.15): $I_{HV} = I_{D0} + \delta/(\delta-1)/(n+1) = 100 + I_a/9$, for $\delta = 5$

P-type (13.25): $I_{HV} = I_{D0} + I_a + I_R$

N-type (13.24): $I_{HV} = I_{D0} + I_k + I_R$

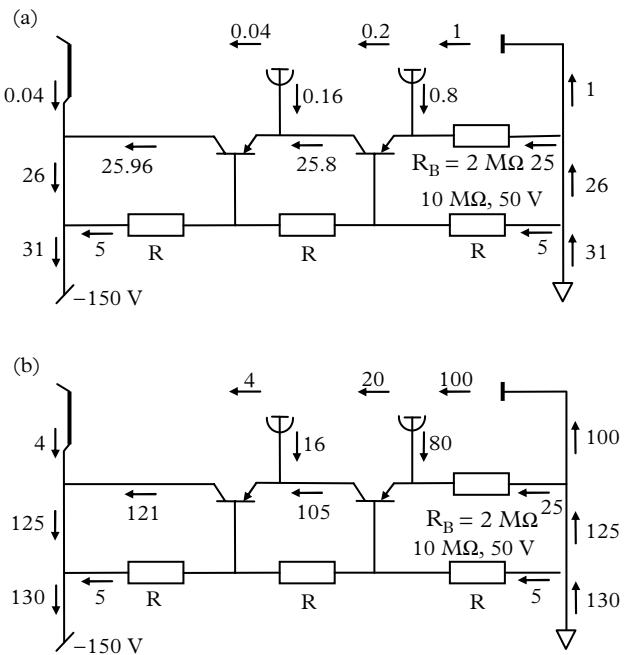
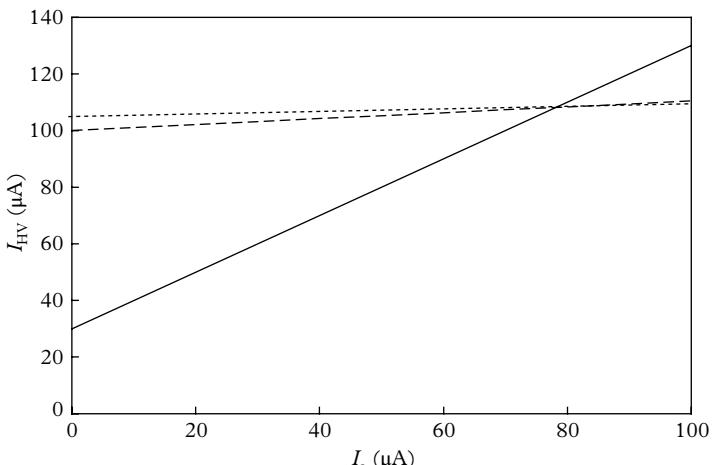


Fig. 13.18. P-type divider draws 30 μA standing current from the supply in the absence of anode current. Note the divider resistors, R , do not enter into the dynamics of the circuit. The significant feature of this circuit is that it supplies anode current on demand. The ballast current, I_{D0} , is fixed at 25 μA , regardless of I_a .



13

Fig. 13.19. The difference in behaviour of the N-type and the P-type active dividers—confirmed by measurement. There is no theoretical limit to the magnitude of I_a for the P-type but, for the N-type, the circuit folds as I_a approaches I_{D0} . The behaviour of a resistive divider is also shown.

13.6.3 Fully active circuits

This analysis shows that a fully active divider may be realized from either N- or P-type devices. The N-type requires a quiescent current in excess of the anticipated maximum anode current. If zener diodes are incorporated at the front end of a P-type divider, then I_{D0} must be at least $20 \mu\text{A}$ for stable operation. The merit of the P-type active divider is its low power consumption—a highly desirable attribute for battery-operated equipment. Another feature of both types of active dividers is scaling of interdynode voltages with HV , analogous to a resistive divider. Typical P-type and N-type MOSFET dividers, protected by 12 V zener diodes, are shown in Fig. 13.20.

The original active dividers reported in the technical literature included signal diodes to provide protection, in one direction only, although some users omitted them altogether. In addition to series decoupling on four stages, parallel decoupling with 1 nF capacitors directly to ground is recommended on the last four

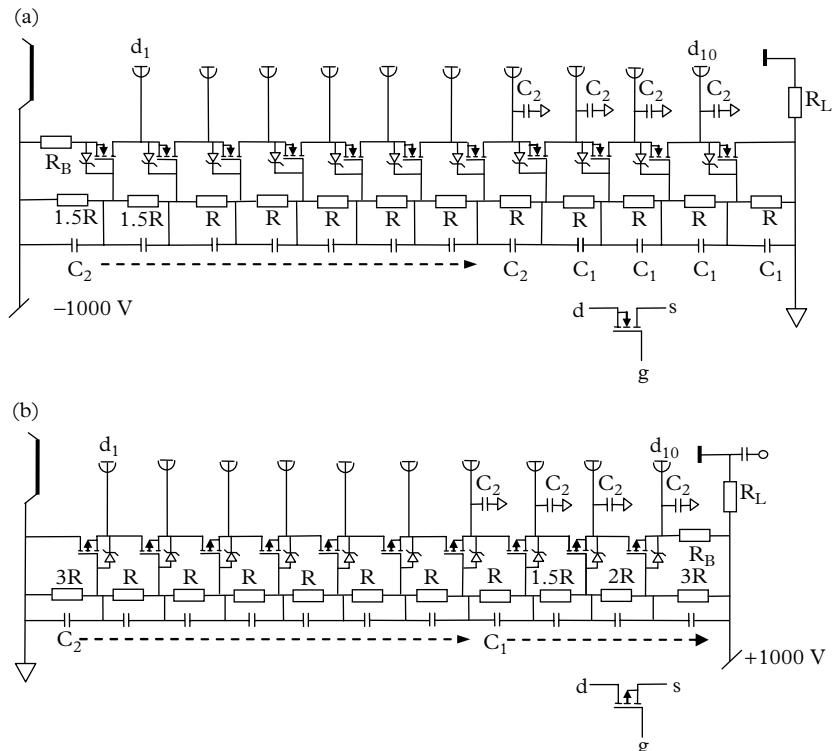


Fig. 13.20. (a) Fully active ten-stage N-type divider; C_1 is typically 10 nF but 1 nF is sufficient for C_2 . (b) Fully active P-type divider. These circuits may be operated with positive or negative HV .

stages. Decoupling dynode resistors is also recommended and in this case 1 nF is sufficient. The circuit illustrated in Fig. 13.20(b) is a tapered divider, recognized through the choice of multiples of R at the back end. The $k - d_1$ voltage is also increased by selecting 3R for the appropriate divider biasing resistor. The parallel string of 10 M Ω resistors plays no part in the dynamics of the divider; this follows because MOSFETs do not draw gate current. The author has found that in practice these divider resistors may be set up to 20 M Ω . Higher resistor values may generate 1/f noise—a form of low-frequency instability. Kerns (1977) draws attention to an inherent safety feature of the N-type configuration, since it is not possible to draw more than the standing current, I_{D0} , from the anode, whereas a P-type turns on harder, with increasing anode current. Thus, P-type dividers present a particular threat to PMT survival, in the event of accidental exposure to bright light.

The key difference between the two configurations lies in setting I_{D0} . In the N-type, the quiescent current is set above the maximum anode current envisaged, while the P-type takes current on demand and, consequently, I_{D0} may be set as low as a few microamps. The dividers shown in Fig. 13.20 are typical practical circuits that include capacitors and protective zener diodes. Capacitors are added to decouple fast and high-amplitude transients—a combination of both series and parallel connections is recommended at the back end. Signal diodes, added across the emitter-base connections, feature in the earlier designs but, nowadays, there is a preference for choosing inactive zener diodes to protect against fast transients of either polarity. However, zener diodes with inherent self-capacitance approaching 100 pF readily transmit such signals to the divider string resistors. This may cause the instability mentioned previously, but the remedy lies in the decoupling arrangement shown in Fig. 13.20. To avoid electrical breakdown, it may be necessary to include additional FETs between those sections where higher interstage voltages are required, for example, $k - d_1$, and $d_{10} - \text{anode}$. This is illustrated in Fig. 13.20(a), where two MOSFET stages supply $v(k - d_1)$.

Bipolar transistors draw base current in response to an input signal, and hence there is always a degree of divider loading with these devices. Nevertheless, an improvement in isolation, compared with that for resistive dividers, approaching h_{fe} is feasible. In contrast, the isolation factor with a MOSFET divider, because of its exceptionally high input impedance, far exceeds that attainable from a bipolar transistor variant. It is therefore feasible to operate with divider resistors of up to 20 M Ω per stage with MOSFETs and, for bipolar transistors, up to about 1 M Ω is generally permissible.

Two model calculations are presented in Table 13.1 for the circuits shown in Fig. 13.20; I_{D0} is assumed to be 100 μA for the N-type, and 10 μA for the P-type.

The resistor labelled R_B , in Fig. 13.15 onwards, sets the standing current in the active string—it is known as a ‘ballast resistor’. In a fully active P-channel divider, R_B is located adjacent to the anode and, for an N-channel circuit, it is near the cathode.

Table 13.1 An illustration of the design procedure for the circuits shown in Fig. 13.20. Recalculation of R will be necessary to achieve the required I_{D0} if the operating voltage turns out to be significantly different from 1000 V.

	N-type	P-type
Total resistance	$13 \times R$	$16.5 \times R$
HV (volts)	1000	1000
$V(R)$, interstage voltage	77	61
$V(R_B)$ (volts)	1.5×77	3×61
I_{D0} (μA)	100	10
R_B ($M\Omega$)	1.15	18.3
R ($M\Omega$)	0.77	6.1

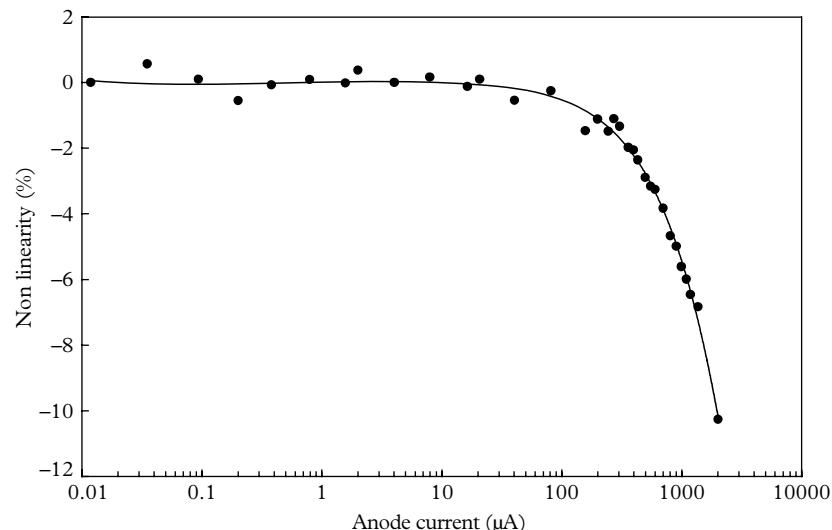


Fig. 13.21. Measured current non-linearity in a 51 mm linear focussed PMT, type ET 9266, operated in a P-channel, fully active divider. An error of measurement of $\pm 1\%$ can be gauged from the scatter in experimental points.

13.6.4 Gain linearity

An example of DC-gain non-linearity is shown in Fig. 13.21 for a 51 mm, fully active P-channel divider operated at 10^6 gain. According to the trend line, linear performance at the 1% level extends to 200 μA , and to 1 mA at the 5% level. The ballast resistor, R_B , is 1.2 $M\Omega$ and all others in this divider are 8 $M\Omega$. Note that

measured non-linearity includes contributions from the PMT and from the divider, but separating these contributions is non-trivial.

13.6.5 Partially active circuits

As the description implies, these circuits include only three or four active devices. Early products predate the availability of suitable FETs, and the reasons for adopting partially active dividers is reduced component count, and hence added reliability. The earliest reported transistorized dividers were all of the partially active type: the back end included, typically, up to four transistors, and the front end consisted of a single series of resistors. Surface mount components and printed circuits were in their infancy in the 1960s and unable to provide the density of tracks that is nowadays commonplace. Component count was therefore a critical consideration with a partially active divider requiring approximately one-third fewer components than the full versions. Another concern, previously mentioned, may have been one of reliability, or perceived reliability, of passive devices over active ones. This does not seem to be of current concern. The author's experience is that properly protected active circuits are as reliable as their resistor counterparts.

A partially active N-type divider using bipolar transistors or MOSFETs is shown in Fig. 13.22. In this example, the currents through the transistors, and that flowing in the d_7 to anode resistor string, are equal. This is the default option but the calculations presented in this section may be repeated by adopting other current ratios. We have, for this particular design,

$$I_{D0} \times 7 \times 150 \text{ k}\Omega + \frac{1}{2}I_{D0} \times 4 \times 300 \text{ k}\Omega = 1000 \text{ V},$$

from which it follows that $I_{D0} = 800 \mu\text{A}$, corresponding to a 90.9 V interdynode voltage. It may be acceptable to reduce this standing current in line with the anticipated maximum anode current.

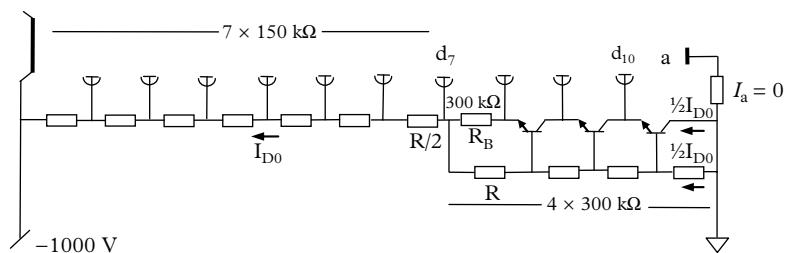


Fig. 13.22. An N-type divider with a quiescent current of $606 \mu\text{A}$. Note that the seven resistors between k and d_7 are common and equal to $R/2$; similarly, those between d_7 and the anode are R . Capacitors and protective diodes have been omitted for clarity.

The analysis in the previous example refers to a given design, whereas a method for determining R_B , to attain any specified standing current through the active string, is also desirable. It follows for a ten-stage PMT with I_{D0} chosen as 100 μA , for example, that

$$R_B = \frac{1000}{11 \times \frac{1}{2}I_{D0}} = 1.8 \text{ M}\Omega,$$

$$R = 1.8 \text{ M}\Omega,$$

$$\frac{1}{2}R = 0.9 \text{ M}\Omega.$$

The four, back-end resistors are also set to 1.8 $\text{M}\Omega$, to obtain the same split.

A more efficient design is realized in Fig. 13.23 by reducing the standing current in the back-end biasing resistors by choosing $R = 10 \text{ M}\Omega$.

With 90.9 interdynode volts per stage, and $R_B = 300 \text{ k}\Omega$,

$$I_{D0} = \frac{90.9}{300 \times 10^3} = 303 \mu\text{A},$$

$$I_R = \frac{90.9}{10 \text{ M}\Omega} = 9.09 \mu\text{A},$$

$$I_R + I_{D0} = 312.09 \mu\text{A},$$

$$R_1 = \frac{90.9}{312.09} = 291.3 \text{ k}\Omega.$$

These currents are shown in Fig. 13.23(b), suitably rounded. If $R \gg R_B$, as in this case, the composite currents may be deduced by inspection. The divider ceases to operate once I_a approaches I_{D0} , driving I_D to zero. Note that the dynode currents in Fig. 13.23(b) injected into the FET string, are 80, 16, 3.2, and 0.64 μA and rapidly sum to 100 μA . The current in d_7 is only 0.67% of the anode current—an indication and measure of the expected gain stability. The maximum anode current that may be drawn is 303 μA , corresponding to $I_D = 0$.

13.6.6 Cockcroft–Walton dividers

The diode–capacitor network shown in Fig. 13.24, the eponymous Cockcroft–Walton (CW) generator, was devised to produce ultra-high DC voltages for pioneering work in splitting the atom. The operating principle is well known and needs no further discussion. Most commercial HV supplies include a CW stack with typically four step-up stages, each of which provides a voltage magnification of 2. These ‘incidental’ voltages are unavailable to the user in a conventional single-output HV supply. However, where PMTs are concerned, they present a way of realizing a uniform voltage divider free from resistors, known variously as a power base or HV base. Voltage dividers based on the CW principle are the most power efficient of all conventional configurations and they are ideally suited to battery-operated equipment, such as luminometers. Figure 13.24 illustrates the operational principle of a feedback-controlled oscillator, operating at a fixed

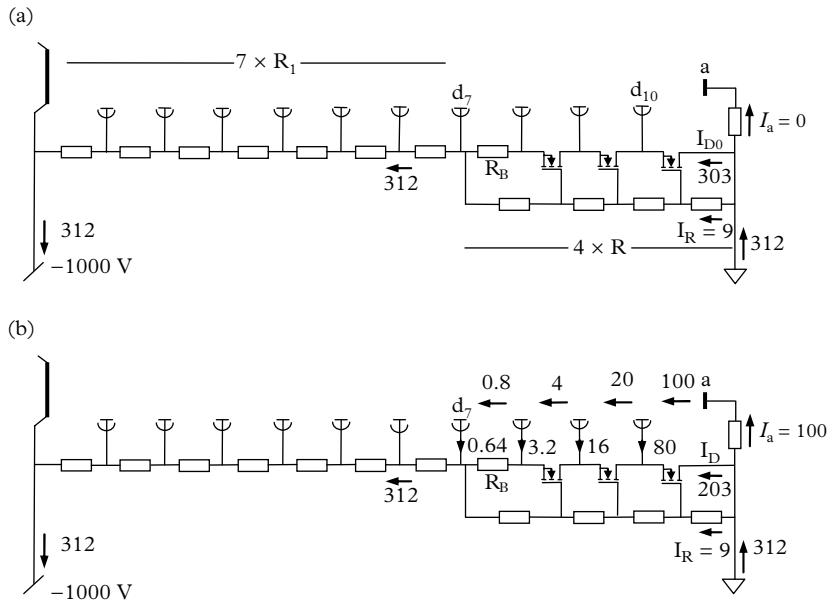


Fig. 13.23. The flow of current in a partially active N-channel divider is shown under (a) quiescent, $I_a = 0$, and (b) signal conditions, $I_a = 100 \mu\text{A}$. Stage gains are taken as $\delta = 5$. Currents are in microamps.

frequency within the range of 100 to 300 kHz. Commercial products require 6 or 12 V input, with power consumption as low as 0.1 W. The oscillator, CW string, and the PMT socket are generally enclosed in a small cylindrical can, directly coupled to the PMT. The can is usually the same diameter as the PMT, and thus well suited for constructing a detector assembly. CW units are capable of maintaining stable gain for anode currents in excess of 200 μA . One limitation of these power supplies is the lack of flexibility in assigning interdynode voltages: each section of the CW divider provides the same voltage step and to achieve twice this, for example at the front end, requires two stages, as shown in Fig. 13.24. The relatively high frequency of the oscillator generates ripple, although the better-designed units achieve less than 1 mV across a 10 k Ω anode load. Ripple increases in proportion to the HV and tends to be worse for positive HV, because of the proximity of the anode terminal to the oscillator. Ripple has the same phase at each dynode with respect to the anode, but it increases in amplitude as the active end is approached. Consequently, power supplies operated with the cathode negative have the lowest anode ripple.

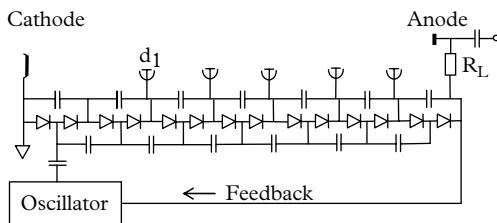


Fig. 13.24. A CW power supply, for positive HV, uses the step-up principle to generate a linear and contiguous set of potentials, for direct connection to the dynodes. Twice the common voltage step has been applied between the cathode and the first dynode, in this example.

13.7 Voltage dividers for pulsed operation

13.7.1 Decoupling capacitors

There are many pulsed applications in which the signal, although short-lived, may attain a peak anode current of hundreds of millamps. Consequently, from Fig. 13.5, the current injected into the last dynode will be reduced only by $(\delta - 1)/\delta$ times the anode signal; the penultimate dynode signal is reduced by a further factor of δ , and so on at each preceding dynode. These transitory current levels are considerably higher than is permissible for continuous operation, but it is possible to comply with manufacturers' recommendations by restricting the rate of events, ensuring that the mean current remains below the maximum permitted. The standing current in a resistive divider is generally less than 1.0 mA, to avoid heat dissipation and, clearly, the injection of transitory currents of around 100 mA will temporarily disturb the divider potentials. However, the addition of decoupling capacitors smooths these current transients, thereby maintaining constant gain over the course of every pulse. The time constants, τ , associated with the dynodes are complex but they are of the order of RC . For example, τ is 3 ms for $R = 300 \text{ k}\Omega$, and $C = 10 \text{ nF}$. Voltage levels are therefore perturbed for ~ 3 ms during and following a current pulse. Decoupling capacitors may be connected between dynodes (known as the series connection) or they may be connected between each dynode and ground (parallel connection). For the moment, we assume series-connected capacitors, leaving discussion of the merits of the two configurations to 13.7.3. Deciding on the capacitance required is straightforward and based on the expression

$$\Delta Q = C\Delta v. \quad (13.26)$$

The charge supplied by the last dynode, ΔQ , is essentially the same as that received by the anode; Δv is the resulting change in the potential of d_n , initially at an interdynode voltage of v (typically of the order of 100 V). Hence, decoupling may be regarded as effective, provided that $\Delta v < 0.1 \text{ V}$, say. Calculating the value of

C required in practice is not always straightforward because at the start the user may have no knowledge concerning the magnitude of ΔQ . However, it is usually feasible to make a guess based on the details of the application—if not, choose 10 nF.

We consider a representative application in scintillation counters. As an example, a fast plastic scintillator, BC104, is to be viewed by a PMT with a requirement to cover a wide dynamic range. There is sufficient information here to decide on the capacitor value. All fast PMTs will show the effects of non-linearity for pulse heights in excess of about 100 mA, so it is sensible to use this as the design parameter for the choice of capacitor. Given that the scintillator decay time, τ , is ~ 2 ns, we can estimate the order of magnitude of ΔQ as follows:

$$\Delta Q = i_{\text{peak}}\tau = 10^{-1} \times 2 \times 10^{-9} = 2 \times 10^{-10} \text{ coulombs.} \quad (13.27)$$

The capacitor required to ensure that $\Delta\nu$ is less than 100 mV is

$$C = \Delta Q / \Delta\nu = 2 \times 10^{-10} / 0.1 = 2 \text{ nF.} \quad (13.28)$$

The capacitor on the penultimate dynode can be reduced by a factor δ , for the reasons previously outlined. Additional scaling applies to the capacitance values for other dynodes. It is usually sufficient, based on years of user experience, to decouple three dynodes or, at the most, four in special cases, with the earlier dynodes relying on stray capacitance for their decoupling. Inadequate decoupling is seldom the reason for poor linearity of response—*intrinsic* PMT behaviour or excessive gain is a likely cause.

In a further example, the capacitance required for the detection of 3 MeV gamma rays with a NaI(Tl) detector is determined. The PMT gain is taken as 10^5 , and the maximum count rate is of the order of 10 kHz. Assume an HV of 1000 V and that $\Delta\nu$ must be less than 0.1 V. A good-quality NaI(Tl) crystal yields 10 pe/keV of ionizing radiation with a decay time constant, τ , of 240 ns. The charge generated at the anode by a 3 MeV gamma is

$$Q = \gamma q e g = 0.48 \text{ nanocoulombs,}$$

where $\gamma = 3000 \text{ keV}$, $q = 10 \text{ pe/keV}$, $e = 1.6 \times 10^{-19} \text{ coulombs}$, and $g = 10^5$.

Taking $\Delta\nu = 0.1$, as before, the capacitor value for the last stage is

$$C = \Delta Q / \Delta\nu = 4.8 \text{ nF.}$$

In practice, 5 nF would be selected for the ultimate dynode, with suitable scaling for the preceding three. For convenience, and to reduce inventory, it is usual to choose a common capacitor value throughout. Note a minimum of four stages should be decoupled in this case, although this advice is not always followed. The mean current at a count rate of $n = 10$ kHz is

$$I_a = nQ = 10^4 \times 0.48 \times 10^{-9} = 4.8 \mu\text{A.}$$

If the acceptable shift in gain, at the highest count rate, is required to be less than 1 %, then, by rearranging (13.21) and assuming $\beta = 9$, we obtain

$$R = \frac{(\Delta g/g)V}{\beta I_a} = \frac{0.01 \times 1000}{9 \times 4.8 \times 10^{-6}} = 230 \text{ k}\Omega.$$

A ten-stage PMT is a likely choice for this application because of the modest gain requirement. If the PMT is operated at ~ 1000 V, in a uniform divider with $R = 230 \text{ k}\Omega$, the required divider current will be in the region of 0.40 mA, and sufficiently low to avoid excessive heat generation. Note that the rate effect is unaffected by the addition or omission of capacitors—their presence improves pulse height linearity only.

13.7.2 Pin connections

Two sets of pin configurations (pin-out) currently in use date back to the first commercially available PMTs. Drawing on sewing technology, the two types are referred to as ‘tacking’ and ‘cross-stitch’. In the tacking configuration, the pin connections, viewed from below, progress in order and clockwise. For a side window nine-stage PMT, the sequence is $d_1, d_2, \dots, d_8, d_9, a, k$. Those following the cross-stitch pattern are configured thus: $d_1, d_3, d_5, d_7, d_9, a, d_{10}, d_8, d_6, d_4, d_2, k$. Pin-connection diagrams, shown in Fig. 13.25, also show internal connections (i.e. include internal connection pins (ic pins)). Although they terminate on a socket pin, they must not be used in the divider circuit construction. Certain PMTs include pins that are not connected (nc pins); these may be used for external mechanical purposes and star connections, given sufficient care. A word of caution: manufacturers exert their right to use all of the ic and nc pins for their own purposes.

The reason for two configurations is as follows: side window dynodes are located on a circular ceramic plate (see 2.6), and electrical access to them is only possible via one of the ceramic supporting structures. The optimal choice of pin configuration is one of mimicking the disposition of the dynodes, thereby avoiding cross over in the wiring. Wires may be attached at many points around the circumference of the supporting flange of a Venetian blind dynode; consequently, PMTs of this type may adopt cross-stitch or tacking in their pin configuration. Because of the dynode layout, linear focussed dynode stacks are compatible with the cross-stitch configuration. The entire side window series of PMTs offered by Hamamatsu Photonics, covering $\frac{1}{2}$ to 1" in size, follow the sequential tacking pattern, with the cathode adjacent to the anode (Fig. 13.25(a)). This close proximity has implications for power supply ripple, as discussed in 13.8.1.

13.7.3 Decoupling options

Three decoupling networks are shown in Fig. 13.26. Series decoupling (a) facilitates ease of component layout and accepts the lowest capacitor voltage rating. Parallel decoupling (b) is preferred where fast pulses are involved because of direct ac grounding. Option (c) is sometimes used for fast-pulse work, including photon

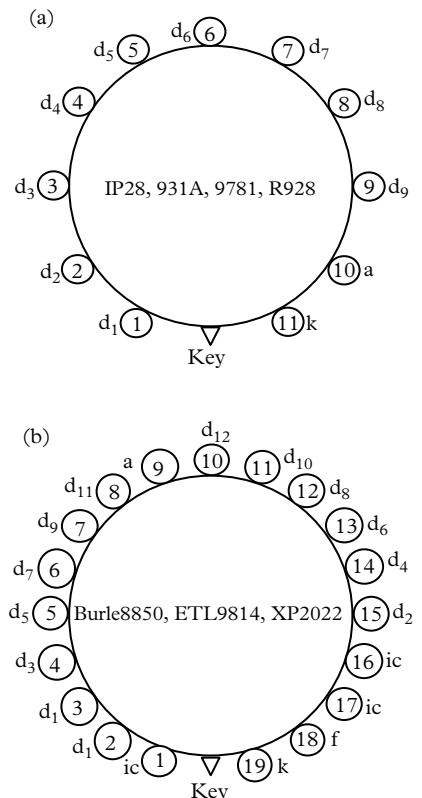


Fig. 13.25. Pin connections for (a) side window and certain Venetian blind PMTs, and (b) fast PMTs. Internal connections are designated ‘ic’. Wiring of the type in (a) can result in short lead lengths for both capacitors and resistors.

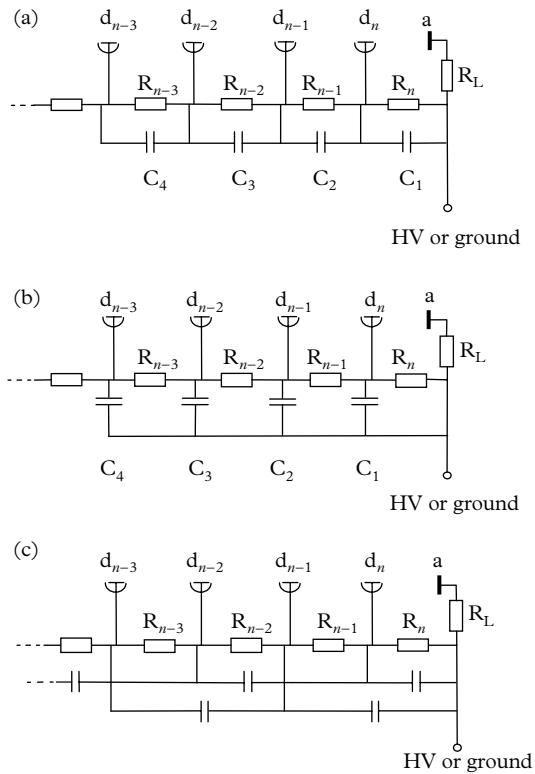


Fig. 13.26. Dynode decoupling: (a) series dynode decoupling; (b) parallel dynode decoupling; (c) series decoupling of dynode pairs.

counting, for PMTs with cross-stitch wiring. Historically, this aided wired component layout, since capacitors follow the tacked arrangement in this configuration. With the advent of surface mount and multi-layer circuit boards, this construction is now of marginal benefit.

13.7.4 Switch-on transients

Series-connected capacitors switch on negative HV faster than parallel versions, as can be gauged from the simulation of Fig. 13.27 for a high-voltage step input. The effect of decoupling further stages is a faster response in the series configuration, and a slower one in the parallel case. The transient at the anode may be of concern where sensitive electronics circuitry is included, particularly with positive HV. It is prudent in this instance to incorporate a basic RC filter to dampen the rise time, as shown in Fig. 13.28(c). Fortunately, most power supplies have a slow start, and the switch-on transient is thus less severe.

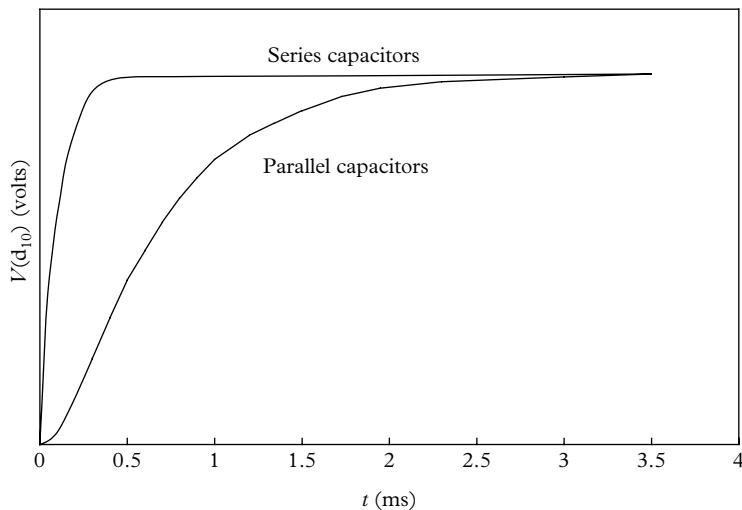


Fig. 13.27. A simulation of the switch-on characteristic on the last dynode of a ten-stage PMT operated with negative HV.

13.8 Methods and circuits

PMTs may be operated with positive or negative polarity, as required by the application. The use of positive HV automatically sets the photocathode to ground potential—this leads to optimal stability of operation. Negative HV is the obvious choice for making DC measurements (analogue operation) and is optional in pulsed applications. The schematics shown in Fig. 13.28(a), (b), and (c) illustrate the complexity of operation with positive HV, while (d) refers to a negative supply. A coaxial power lead, decoupled by C_1 , delivers HV to the PMT. Positive operation is influenced by the juxtaposition of the HV input to the anode—the inclusion of C_1 is essential, regardless of whether a coaxial HV lead is used, or a wire connection is made; R_s is a safety resistor that prevents the output from charging to anode potential, on disconnection of the PMT output. It is advisable to adopt this resistor when using positive or negative supplies.

13.8.1 Power supply polarity and ripple

The behaviour of voltage-divider circuits, in the presence of ripple, is frequency dependent because of the inverse relationship between capacitance and impedance. For example, a capacitor of 10 nF has an impedance of 159 Ω at 100 kHz but 318 k Ω at 50 Hz. Consequently, effective decoupling is difficult to achieve for



Fig. 13.28. The distribution of ripple in standard biasing arrangements. Typical component values are $R_L = 100 \text{ k}\Omega$; $R_s = 10 \text{ M}\Omega$; $R_f = 100 \text{ k}\Omega$; and $R = 330 \text{ k}\Omega$. Arrows indicate ripple paths; (a) and (b) represent the basic decoupling configurations; (c) shows a variant with an added pi network; (d) refers to negative HV; and (e) illustrates the contributions from back-end dynodes via stray capacitance.

mains-generated ripple. Capacitors are included in the circuits to restore the signal charge required of the dynodes, and thus to maintain constant gain in the presence of transient events. However, capacitors also affect power supply ripple as seen at the anode, and not necessarily advantageously. Physical size often limits the choice of capacitor, particularly for surface mounted circuits where 10 nF may be desirable but not necessarily available for HV requirements. The effect of power supply ripple can be studied by including a sine wave generator, v_r , in the circuits of Fig. 13.28, with the HV supply switched off.

First, consideration is given to power supplies showing ripple of 100 kHz in the series-connected circuit shown in Fig. 13.28(a). It should be noted that the voltage developed across C_1 is also present in proximity to the anode, and markedly if $R_s \gg R_L$. The decoupled section of dynodes is effectively a short circuit at this frequency, and ripple current propagates down the divider string to the cathode. This ripple modulates the potentials on the dynodes and, if sufficiently severe, it will affect gain and pulse height resolution. The second consideration is the presence of 50 Hz ripple, at which frequency the impedance of the capacitors is comparable to, or may even exceed, that of a typical interdynode resistor. Consequently, the anode detects the full amplitude of ripple present at d_{10} , and the remaining dynodes receive a scaled-down proportion of this ripple. Third, referring to a parallel-connected circuit shown in Fig. 13.28(b), high-frequency ripple at the anode is slightly less than that recorded for the series case; the ripple on the dynodes, however, is highly attenuated because the decoupling capacitors direct ripple to ground. Where 50 Hz ripple is generated, that present on the dynodes is a factor of about 10 lower than for the series version. In extreme cases of ripple, it may be necessary to insert the pi network, comprising R_f , C_1 , and C_2 , as shown in Fig. 13.28(c).

With negative HV, ripple is developed across the dynodes in much the same way as with positive HV (Fig. 13.26(d)). However, the voltage-divider network terminates at ground potential and hence the ripple current is diverted away from the anode. A ripple signal is, however, coupled into the anode through inter-electrode capacitances at the back end of the PMT (this also applies to the positive option), as shown graphically in Fig. 13.28(e). The associated stray capacitances (see Table 13.2) are of the order of 3 to 5 pF and, consequently, the amplitude of this ripple is small, even at 100 kHz . The important role of C_1 is evident in all the figures and this should be noted. It may be necessary, with reference to Fig. 13.28(d), to omit R_s with the use of certain electrometers and charge sensitive preamplifiers. Both these electronic circuits are delicate and easily offset, warranting special care.

Ripple is seldom of concern in detectors where matched coaxial cable is used for signal transmission. Only the configuration shown in Fig. 13.29(c) is sensitive to ripple, via the direct 50Ω connection to the power supply, and so is not recommended. Ripple amplitude at the anode varies linearly with HV and inversely with R_L . It may be feasible to raise HV to escape ripple by taking advantage of the power law dependence of gain. Voltage dividers suitable for fast pulses, typically of nanosecond duration, demand additional components and particular attention to wiring—these topics are covered in 13.8.8.

Table 13.2 Measured capacitances between various electrodes. Capacitances (C) increase by about 0.5 pF by adding a socket.

Type	area (cm ²)	C (pF)
Interdynode		
Venetian blind	10	11
Linear focussed, 50 mm	2	3
Linear focussed, 30 mm	1	2
Circular focussed, 25 mm	0.5	1.5
Anode-last dynode†		
Venetian blind	5	6
Linear focussed, 50 mm	2	3.5
Linear focussed, 30 mm	1	3
Circular focussed, 25 mm	0.5	1
Anode-all dynodes†		
Venetian blind	5	6
Linear focussed, 50 mm	2	5
Linear focussed, 30 mm	1	5
Circular focussed, 25 mm*	0.5	2

† The anode is a wire over a plane.

13.8.2 Wiring options

There are three mechanical pin-out configurations covering the range of PMTs, although not all are available in every PMT type. Selection must be made from the options quoted in manufacturers' data sheets:

In a 'hard pin base', pins exit directly from the glass envelope to a mating socket. PMTs of this configuration have the shortest overall length.

A 'capped base' is bonded to a PMT and offers robust pins, set in opaque thermoplastic material. This PMT option is longer than the hard-pin version but it provides excellent electrical contact and mechanical support.

A 'flying lead base' has a set of flexible wires suitable for soldering directly to a circuit board, but care must be taken to remove remnant flux. Alternatively, the wires may terminate in a loose cap (unsoldered), for users who wish to make a light tight detector module.

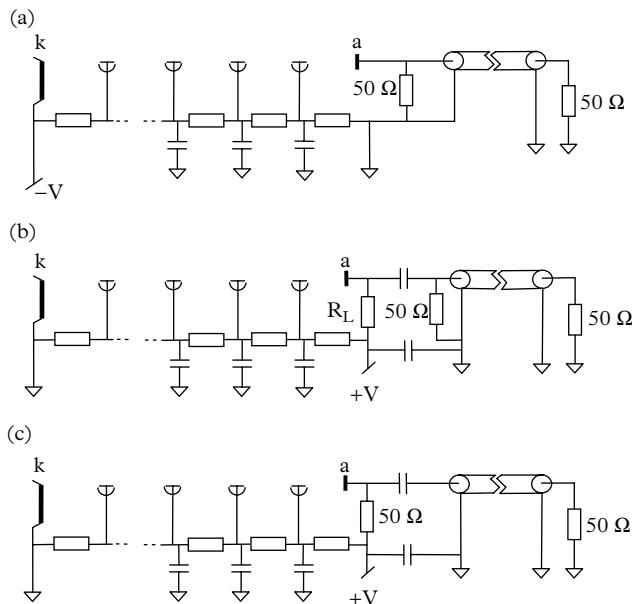


Fig. 13.29. Matched voltage dividers for driving coaxial cable. Configuration (a) is ideal and based on a negative supply; (b) is immune to power supply ripple, provided that $R_L >> 50 \Omega$; (c) is usually unsatisfactory because ripple is conveyed by the 50Ω load directly to the anode.

Voltage divider components, in the form of radial resistors and leadend capacitors, may be attached directly to the socket contacts, although there is invariably a physical layout problem with grounding the braids of HV and signal leads. A printed circuit board, using surface mount components, is the preferred choice for most users. Components may be mounted on both sides of the board, with the inclusion of a ground plane (important for ultra-fast signals). Sockets with the option of PCB contacts are ideal for printed circuit boards. Customers are often surprised to discover that the cost of a single socket and voltage divider can be a significant proportion of the cost of the PMT. This does not apply to significant quantities of boards, for which manufacturers are able to pass on the benefits of scale. Those wiring a divider for the first time will invariably get it wrong—even experienced users of PMTs are not exempt. The choice of positive HV always presents an extra challenge, and users are advised to probe a divider circuit, initially with minimal HV, before inserting the PMT.

13.8.3 Dynode signals

Signal current flows into every dynode of magnitude $I_a(\delta-1)/\delta$ at the last dynode, $I_a(\delta-1)/\delta^2$ at the penultimate dynode, and so on. Dynode signals are

reduced-in-magnitude and inverted versions of the anode signal. A voltage signal can be generated by inserting a load resistor in series with a dynode, as shown in Fig. 13.30. The magnitude of R_{L1} has the same range as an anode load. There are two reasons for deriving a dynode signal from a PMT. First, it makes it possible to extend the dynamic range of linear operation by taking advantage of the reduced gain that applies at the dynodes; gain is reduced by a factor of $\sim\delta^2$ for the examples shown. Signals may suffer from distortion or saturation above a certain level at the anode but, because of their reduced size, they will remain linear at the chosen dynode. This arrangement always outputs two signals, and the user can choose to ignore either of them if not required.

Second, it is possible to achieve both fast timing and pulse height encoding simultaneously. It is customary to use the dynode signal for timing, as there is evidence that this signal is faster than the one at the anode. Either series or parallel connections for the decoupling capacitors may be adopted but an AC earth created by C_5 is essential to maintain a flat baseline. Signals can be picked off any dynode, but the ultimate dynode is generally disfavoured for being too closely coupled to the anode by internal capacitance.

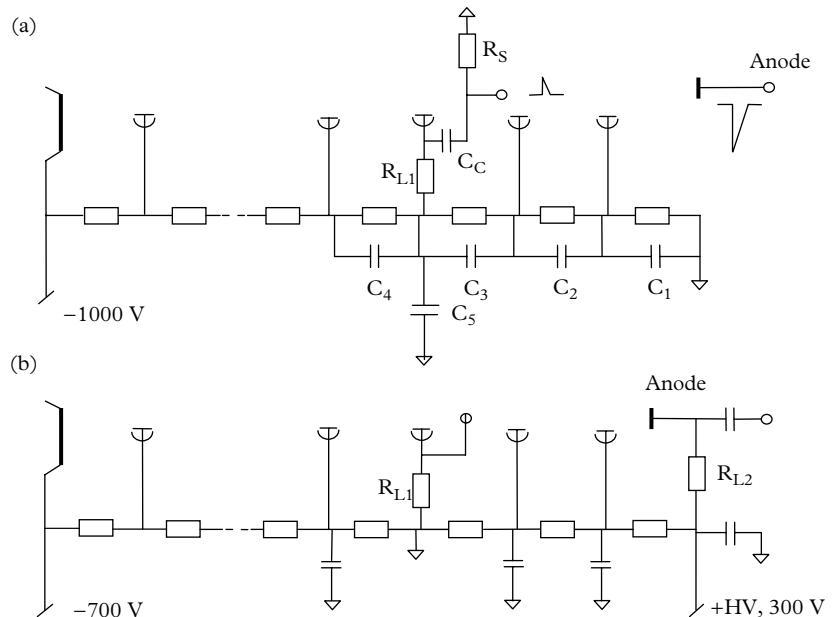


Fig. 13.30. Two ways in which to pick off a dynode signal. In (a), the signal is AC coupled with the anode grounded. For DC coupling, the arrangement with two power supplies in (b) is used. Note the dynode resistor R_{L1} may be omitted in (b), depending on the input impedance of associated electronics.

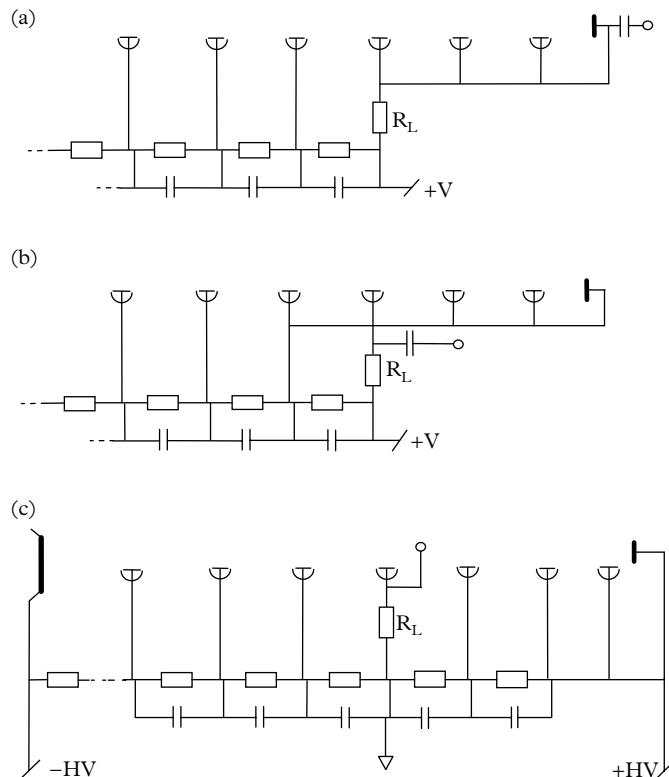


Fig. 13.31. Three configurations for operation with fewer dynodes. (a) Not recommended if linearity and speed of response are critical. (b) The dynode from which the signal is taken is made to resemble an anode, but not very convincingly. (c) Provides best performance but needs two power supplies. Note the similarity between Fig. 13.28(b) and Fig. 13.29(c).

13.8.4 Shorting dynodes

Shorting dynodes is considered when it is discovered that light levels are higher than expected and a given PMT has too many stages. The possible consequences of operating at low gain, and hence low interdynode voltages, are a sluggish time response, gain that is sensitive to small changes in HV, and poor linearity. The best advice is to replace the PMT with one of fewer stages; the alternative is to short a set of consecutive dynodes to the anode. This may prove to be the only option, since the available choice of PMTs with fewer than six stages is very limited.

The configuration in Fig. 13.31(a) is customarily adopted, although it is the least desirable way of reducing the number of active dynodes, regarding linearity and speed of response. The arrangement in Fig. 13.31(b) attempts to mimic a

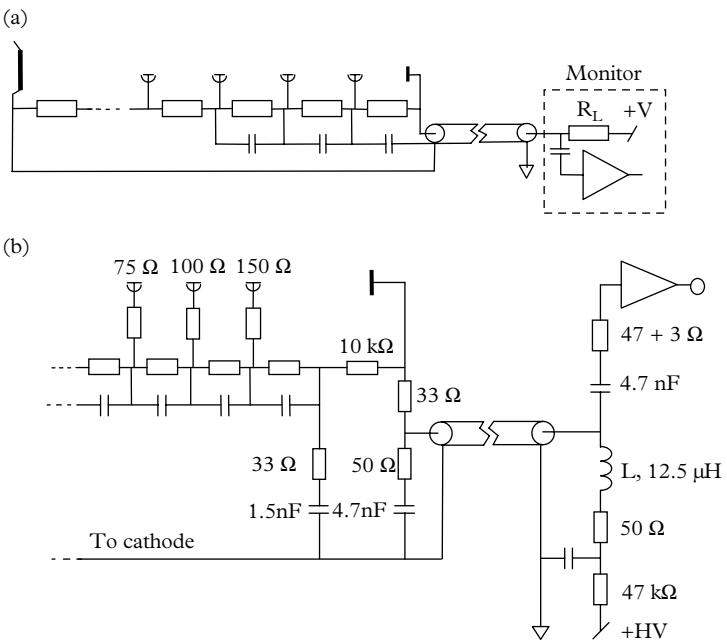


Fig. 13.32. (a) Basic single-cable divider, used mainly for portable equipment. An anode series resistor may be included but it is unnecessary. (b) A critically tuned version developed by Brigatti *et al.* (2005).

reflex anode and results in better performance than the configuration shown in Fig. 13.31(a). In the configuration shown in Fig. 13.31(c), the signal is allowed to pass beyond the pick-off point where normal biasing conditions apply. This prevents the early onset of space charge effects, which soon become evident with the configurations shown in Fig. 13.31(a) and Fig. 13.31(b). Note that the signal in the configurations shown in Fig. 13.31(a) and (b) is negative while, in the configuration shown in Fig. 13.31(c), it is positive.

13.8.5 Single cable dividers

PMT-based instrumentation, such as handheld radiation probes, favours the adoption of a single coaxial cable for delivering positive HV, and for transmitting the signal to a control unit. The primary reason for its use in radiation monitors is simply the practical advantage of a single cable. The schematic is shown in Fig. 13.32(a) but note it is for positive HV only. An advanced form of this divider is described by Lagomarsino and Testera (1999) and Brigatti *et al.* (2005) for use in a massive astrophysics application (Borexino). The adoption of this technique halves the cost of cables, representing a considerable saving in this instance. The impedance of coaxial cable is strictly matched with a pure resistor, typically $50\ \Omega$,

but this is compromised once a coupling capacitor is introduced. The interesting feature of this divider is the matching obtained by equating time constants L/R and RC to achieve $R = (L/C)^{1/2} = (12.5 \mu\text{H}/5 \text{nF})^{1/2} = 50 \Omega$. It follows from circuit analysis that this combination of components has an impedance of 50Ω at all frequencies, thus eliminating reflections down an otherwise mismatched line. Signal preservation is also evident at the anode end of the line through the use of 33Ω resistors, together with the addition of graded damping resistors at 75 , 100 , and 150Ω .

13.8.6 Gain adjustment

The easiest method for gain adjustment of a PMT is to vary the HV. Where several PMTs share the same supply, voltage dropping resistors may be used for individual control. The power supply load regulation, together with the combined divider current, ultimately limits the number of PMTs that can be served by a single supply.

A second method can be adopted by inserting a variable resistor between two stages, say, d_5 and d_3 , with d_4 connected to the wiper. The potentiometer resistance is $2R$, where R is the common resistor value, with the advantage that the load presented to the supply remains independent of the potentiometer setting. If the wiper is off-centre by $\Delta\nu$ volts, the combined gain of the two stages, using (13.1), is

$$\delta_n \times \delta_{n+1} = a(\nu + \Delta\nu)^\alpha \times a(\nu - \Delta\nu)^\alpha = a^2[\nu^2 - (\Delta\nu)^2]^\alpha. \quad (13.29)$$

Gain is a maximum when the wiper is set to the mid-point, and movement in either direction has the effect of reducing gain over a usable range of 2:1. There is a danger with settings close to either extremity that linearity and speed of response will suffer; in general, the potentiometer should be positioned between d_3 and d_5 , or thereabouts.

13.8.7 Equivalent circuit at high frequencies

The equivalent circuit of a PMT consists of a multiplicity of inductors, mutual inductors, and capacitors interconnected in a complex network: all fast PMTs demand respect for high-frequency techniques. A reflex anode is the preferred construction, consisting of a mesh positioned between the last two dynodes, which themselves connect to pins adjacent to the anode pin. This represents an attempt to achieve a transmission line, albeit a poor one, between the anode mesh and its pin connection. A pure transmission line is characterized by a capacitance per unit length, C , and an inductance per unit length, L . Expressions for various types of transmission line may be found in electronics textbooks. Three configurations of particular relevance to PMT signals are shown in Fig. 13.33 and refer to vacuum operation (no dielectric). The inductance per metre, L , and the characteristic impedance, z_0 , follow once C is known, as follows:

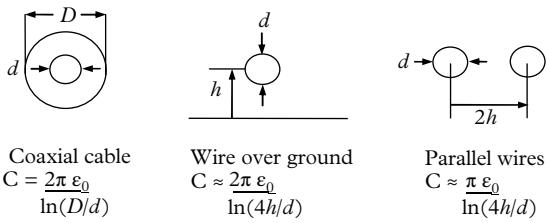


Fig. 13.33. Three types of transmission lines and expressions for capacitance per metre where $\epsilon_0 = 1/(36\pi \times 10^9)$ F/m.

$$L = 1/c^2 C$$

$$z_0 = 1/cC = (L/C)^{1/2}, \text{ where } c = 3 \times 10^8 \text{ m/s}$$

$$T = (LC)^{1/2}, \text{ usually expressed in nanoseconds per metre}$$

A possible configuration for signal transmission in a PMT is that of parallel wires: the principal one is the anode lead, and the other two serve the last and penultimate dynodes. Capacitances, and hence all the other parameters, are insensitive to an exact knowledge of spacing because the parameters d , D , and h enter the formulae logarithmically. Taking $d = 0.7$ mm, and $2h$ as 5 mm leads to $C = 10$ pF/m; it follows that $L = 1.1$ μ H/m, and $z_0 = R_d = 300$ Ω .

An obvious shortcoming of this approach is that it deals with wiring only, although interdynode capacitances far exceed those of the wires alone. Capacitance measurements by the author are given in Table 13.2, taken, obviously, by using non-operating PMTs.

A different approach is to consider the wiring as a tuned circuit, with C values based on Table 13.2, and L taken as 100 nH. Assuming a typical capacitance from Table 13.2 of 5 pF, then

$$f = \frac{1}{2\pi(LC)^{1/2}} = 250 \text{ MHz.} \quad (13.30)$$

The circuit is critically damped when

$$R_d = (L/(2C))^{1/2} = 100 \text{ } \Omega. \quad (13.31)$$

Of the two approaches to understanding the behaviour of a fast PMT, it would appear that the critically damped resonance model is the more plausible one.

13.8.8 Wiring practice

Where the signal is transmitted from the anode pin via coaxial cable, best practice is to use the outer braid as the common earth point, as shown in Fig. 13.34. Matching or ‘damping’ resistors, denoted R_d and inserted in series with the last two or three dynodes, are recommended to suppress ringing on the trailing edge of

pulses. The author's experience is that a value between 30 and 150 Ω will suffice, and the last dynode is responsible for most ringing. These resistors may be concealed within the cap, if this form of base is selected; otherwise, they are included in the divider board or on the socket. Decoupling the HV cable with C_1 directly to the common ground is essential; alternatively, following Kerns (1977), a capacitor in series with a resistor of characteristic impedance—inevitably, 50 Ω —may be used. The C_2 capacitors, terminating the R_d resistors, need to be low inductance: 1 nF is sufficient for this purpose, and similarly for C_1 . The technique used in time-domain reflectometry is a useful diagnostic tool for locating sources of ringing, and the origin of delayed satellite pulses. This is achieved by changing sequentially the length of the two coaxial cables and observing the effects on an oscilloscope. The schematic shown refers to a flying-lead PMT, but the rules are the same for a hard-pin type, where connections are made to socket contacts. The dominance of 50 Ω input and output impedance for instrumentation undoubtedly explains why this is also the standard for PMTs, although using 93 Ω cables, for example, would generate nearly twice the output voltage, for the same current input. A 93 Ω output impedance is sometimes adopted in charge sensitive preamplifiers.

Figure 13.35 illustrates the consequences of an improperly constructed divider. Clean single pulses such as those in Figs 13.36 and 13.37 can be realized at the anode by following the recommendations embodied in Fig. 13.34. Signal pulse

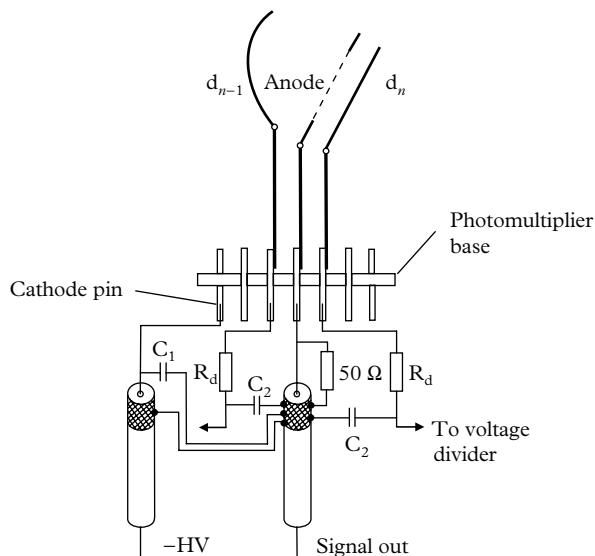


Fig. 13.34. Wiring construction for a flying-lead PMT. Damping resistors, denoted R_d , are connected to the voltage divider, where additional decoupling is added. The use of screened cable for the HV is essential, with the outer braid earthed directly to that of the signal cable—the star point. Arrows show connections to a divider circuit board.

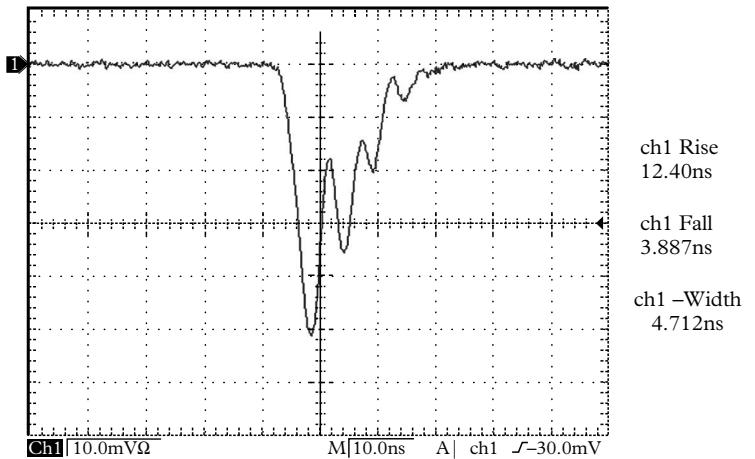


Fig. 13.35. An extreme example as a consequence of poor wiring practice with positive HV and a 9107B PMT. Ringing at 170 MHz is due mainly to inadequate HV supply cable decoupling.

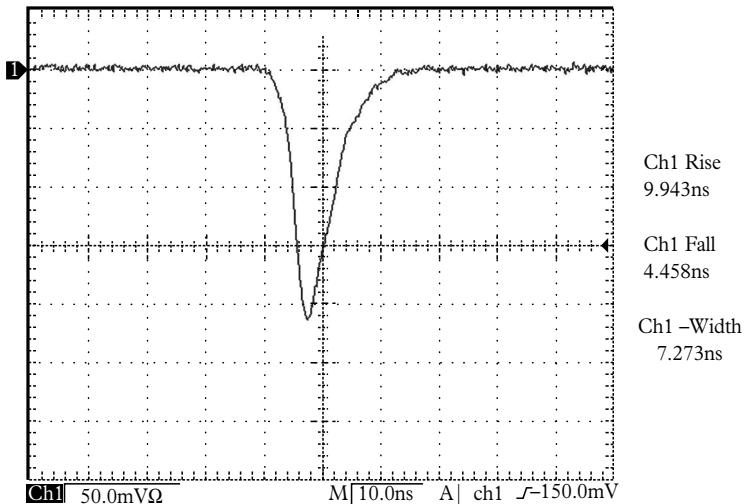


Fig. 13.36. The output pulse obtained after adopting the techniques shown in Fig. 5.32. This is from the same PMT as used in the previous figure.

shapes that may have appeared clean, before the general availability of wide bandwidth sampling oscilloscopes, in reality masked high-frequency oscillations. An important question remains: are structured trailing edges merely ugly, or are they a serious limitation? Applications involving energy encoding (that is,

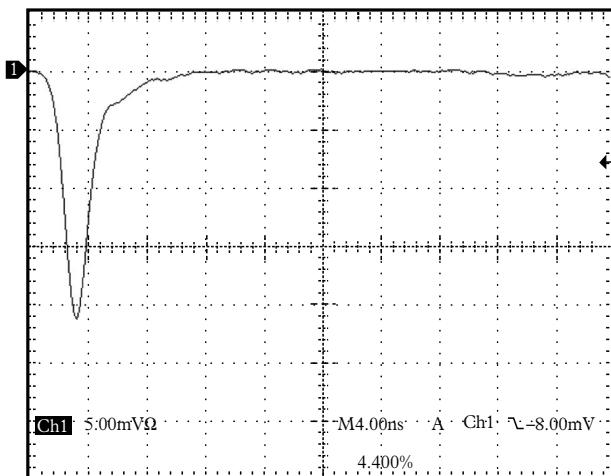


Fig. 13.37. Single-electron pulse for a 9142B, with a leading edge of 1.4 ns and a fwhm of 2 ns, in the same divider construction as that adopted for Fig. 5.31. There is barely discernible ringing but it is masked by the dog leg at the base of the trailing edge. This common feature of fast PMTs originates in the electron optical design, but it can be removed by stub matching (see 8.5).

integration) are arguably unaffected. Discriminators, however, are prone to retrigger on structured signals, and particularly on satellites generated by mismatched cables. PSD is clearly an application that demands clean signals.

13.9 Gating

There is sometimes a requirement to measure transient low-level signals occurring shortly after an intense pulsed-light signal. The intention in these applications is to measure the daughter products of an intense flash, and not the flash itself. Various time-resolved spectroscopic techniques, conducted under laboratory conditions, entail exciting a sample with an intense laser pulse, for example. This reveals phosphorescence and fluorescence emissions over a delayed time interval, spanning up to milliseconds. LIDAR (for *light detection and ranging*) concerns the detection of far-field backscattered light generated from a high-powered pulsed laser; light pulses are launched vertically into the atmosphere at a low repetition rate. Gating is mandatory to suppress intense near-field scattered light from buildings and low-level clouds. Gated PMTs also find application in nuclear physics for observing radiation phenomena following intense X-ray or beta radiation. All PMT gating configurations fall within the two categories illustrated in Fig. 13.38.

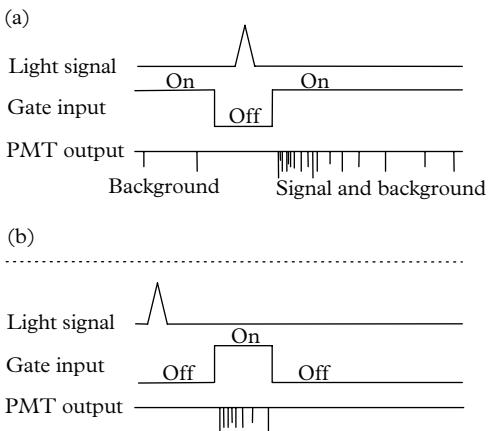


Fig. 13.38. Two options for gating PMTs. (a) The PMT is normally on, producing background signals and switched off while exposed to an intense transitory light signal. Once the initiating light signal is over, the PMT is switched back on, producing signal and background events. (b) The PMT is normally off prior to and during an initiating event. Signal and background are present only during the on-state.

In the absence of gating, or where gating is inadequate, the response of a PMT to the initiating signal may introduce artefacts that interfere with the proposed measurements. These are predominantly signal-induced afterpulses and enhanced PMT background. In addition, voltage-divider biasing voltages may depart significantly from their quiescent settings, to return only after many milliseconds, and PMT gain is consequently unstable over this period. Extreme exposure may result in electrical breakdown, leading to permanent damage to the PMT and any ancillary electronics. The simplest means of avoiding these spurious phenomena is to include mechanical shutters or choppers in the measurement method, but this adaptation is restricted to slow systems.

A better, but more challenging, technique lies in turning the PMT off for the duration of an intense signal. In practice, an electrical means has to be found to reduce the gain during the presence of the initiating light pulse. The extinction coefficient, γ , is the primary specification for gated operation: it is defined as the ratio of the PMT gain under gated conditions to that for normal low-light-level operation. As noted in the review by Wardle (1982), published gating systems tend to relate to a particular type of PMT and they may not transfer to PMTs of different internal structure. In addition to optimizing γ , other performance specifications such as speed of response, signal-to-background enhancement, and switching transients may conflict. It has been proposed that gating could be achieved by pulsing a fine mesh attached to the outside of a PMT window but this does not work, presumably because of the semiconductor nature of photocathodes

It is convenient to consider pulsed gating circuits under six categories:

- (1) mechanical shutters and choppers
- (2) electro-optical shutters
- (3) HV switching
- (4) photocathode
- (5) focus electrode
- (6) dynodes

13.9.1 Electro-optical and mechanical gating

The ideal way to gate a PMT in the presence of an intense light source is to prevent such light reaching the photocathode. Electro-mechanical shutters and choppers offer reasonably high extinction, but they are constrained to millisecond rise and fall times, with associated jitter. Improved performance can be attained by adopting electro-optical devices in place of electro-mechanical ones. Certain crystals and liquids react to an applied electric field by a change in refractive index, polarization, and hence transmission. In the Pockels effect, the refractive index varies linearly with the applied electric field, whereas, in the related birefringence phenomenon, the Kerr effect, transmission is proportional to the square of the electric field. To achieve fast shuttering requires a pulse of 1 to 5 kV with a Pockels cell, as compared with 10 to 20 kV for the Kerr effect. The literature on this subject is rather sparse although the description of a particular LIDAR application may prove useful (Kirillov and Samokhvalov 2014).

13.9.2 HV switching

Wardle (1982) makes reference to four investigations based on switching the HV supply. Of the previously mentioned categories of pulsed gating circuits, De Marco and Penco (1969) have investigated (3), (4), (5), and (6) but the challenging mode is undoubtedly (3), where a delay-line thyratron pulse generator activates the PMT. Thyristors are solid-state, negative-resistance successors to gas-filled thyratrons, which they have replaced in most switching applications. This unconventional voltage divider consists of 16 equal, small-value resistors with a total impedance of $50\ \Omega$, which matches the transmission line that delivers the switch-on pulse. While the gain restoring rise time is 10 ns, the authors abandoned this approach because of pick-up transients observed at the anode. The origin of these may be a mismatch caused by the choice of interdynode capacitances, which also direct fast edges to the anode. Power MOSFET circuits can be configured to switch HVs and currents if properly designed. The challenge when devising a nanosecond switch is charging and discharging the input capacitance of the MOSFET. This capacitance is of the order of 1000 pF and therefore difficult to drive in conventional common drain configurations. However, Baker and Johnson (1992) avoid the problem by adopting two MOSFETS connected in cascode.

13.9.3 Photocathode gating

A common difficulty with photocathode gating schemes is the resistivity of the cathode, and the capacitance between the photocathode layer and neighbouring electrodes. In conventional PMTs, a transparent substrate of platinum or an evaporated aluminium grid, on the vacuum side of the PMT window, is beneficial, but fast switching and recovery are still elusive. This is especially acute in bialkali types (see 2.7). Gating MCP-PMTs is considerably more successful, especially in those devices that include highly conductive S20 and GaAs photocathodes. Photek Ltd supply MCPs in the PMX series with the option of a gating-control module (GM150-20) capable of achieving nanosecond switching. The Hamamatsu R5916U series includes a grid electrode mounted in close proximity to the photocathode. According to the manufacturer, this construction has been superseded by gating the photocathode directly, yielding higher sensitivity and faster switching.

13.9.4 Dynode gating

In theory, any one or more dynodes may be gated, but some perform better than others. This is easily ascertained, for a particular PMT, by varying the potential of a single dynode from its ungated state. Figure 13.39 illustrates gating at the second dynode of an ET 9214 B, using a separate supply, V_{gate} . Multiple dynode gating can be explored using a second independent power supply. Clearly any dynode,

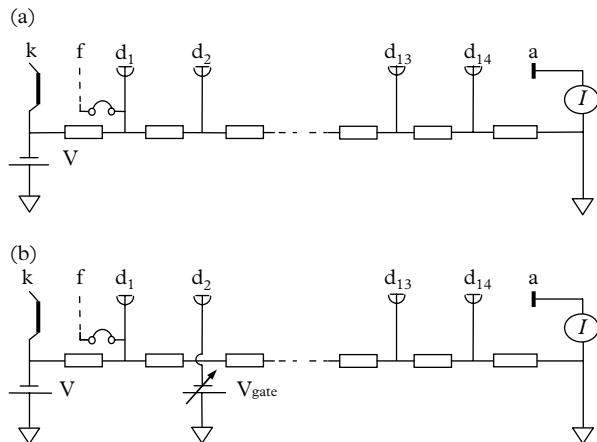


Fig. 13.39. (a) The ungated state of a divider; (b) an adaptation for measuring the efficacy of gating d_2 , for example. The HV supply, V , is fixed at -1400 volts while V_{gate} is varied over the range 0 to $-HV$. Variation in anode current, derived from a continuous LED light source, leads to the extinction coefficient γ as a function of V_{gate} , as shown in Fig. 13.41.

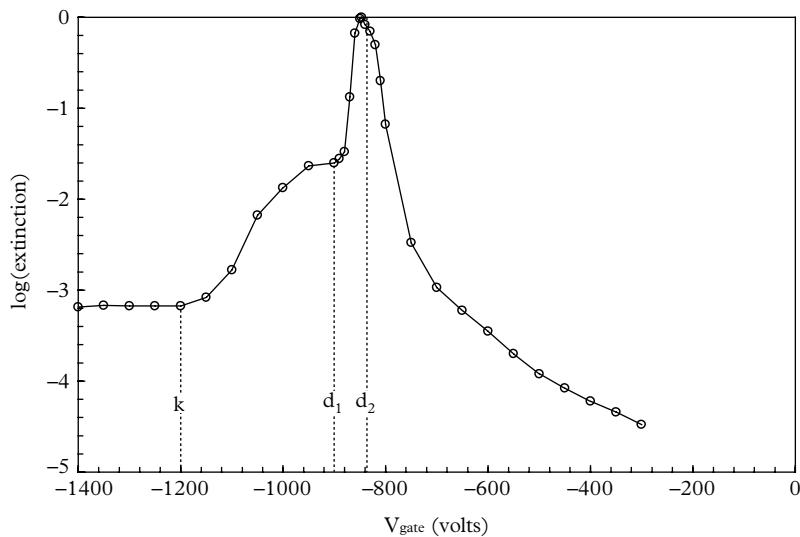


Fig. 13.40. The effect on anode current of applying a gate signal to d_2 .

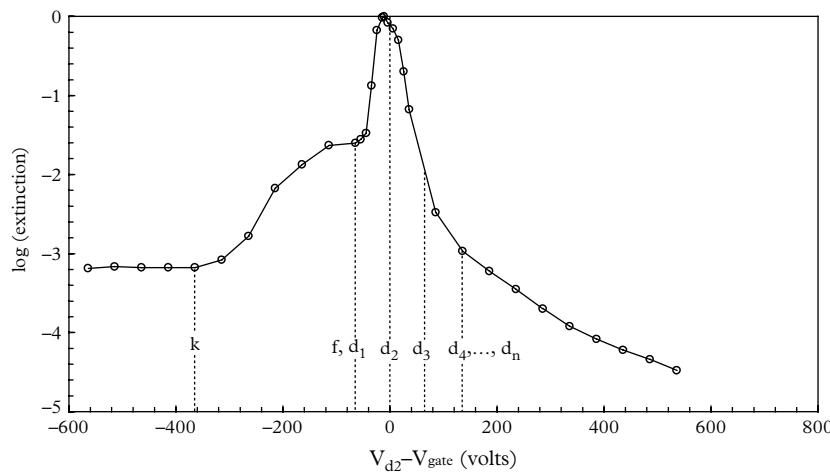


Fig. 13.41. The same data as in Fig. 13.40 but plotted with respect to V_{d2} .

and the focus electrode, may be gated in this way, examples of which are shown in Figs 13.40–13.43.

Such DC gating has no obvious application other than uncovering the dynodes best suited to this conceptual form of gating. The performance under fast pulsed conditions is likely to be different in detail from DC performance but such investigations serve as a necessary test for a given PMT. The abscissa in Fig. 13.40 is V_{gate} but for comparative purposes it is more informative to plot the

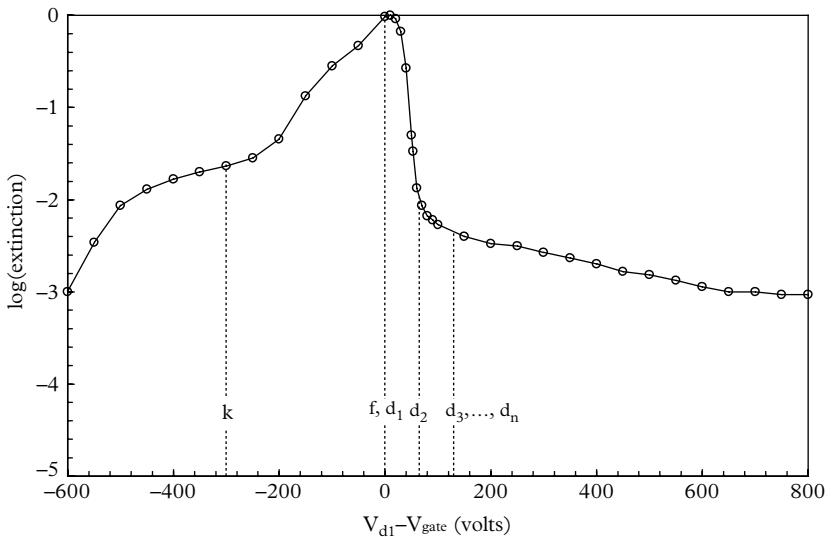


Fig. 13.42. Gating d_1 requires a higher voltage than for other dynodes—not recommended.

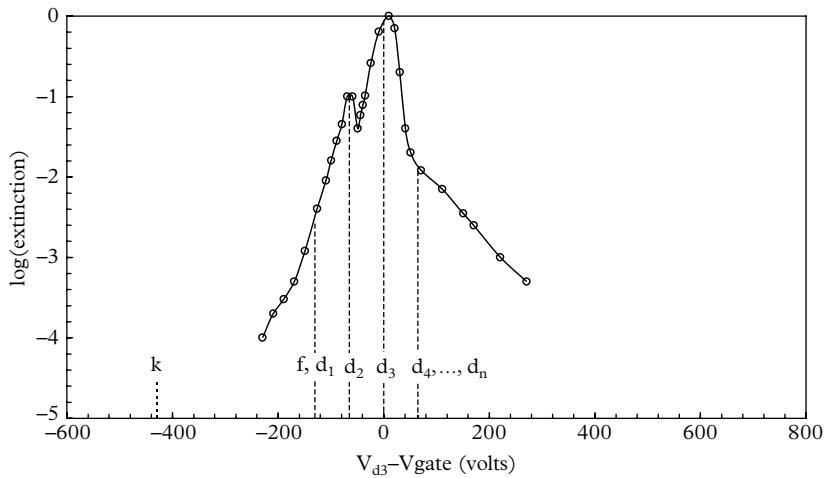


Fig. 13.43. The origin of the artefact at d_2 is unknown. Otherwise, gating d_3 appears satisfactory.

graphs with respect to the dynode concerned (see Fig. 13.41). The overall pattern for gating d_2 is repeated for other dynodes: peak gain corresponds closely to the ungated condition, and gating is most efficient for $V_{d1} < V_{\text{gate}} < V_{d3}$, with γ reaching $\sim 1/40$. Extinction increases either side of these limits but the rate of improvement is modest. The extinction ratio approaches 10^{-3} by gating d_2 positively, but

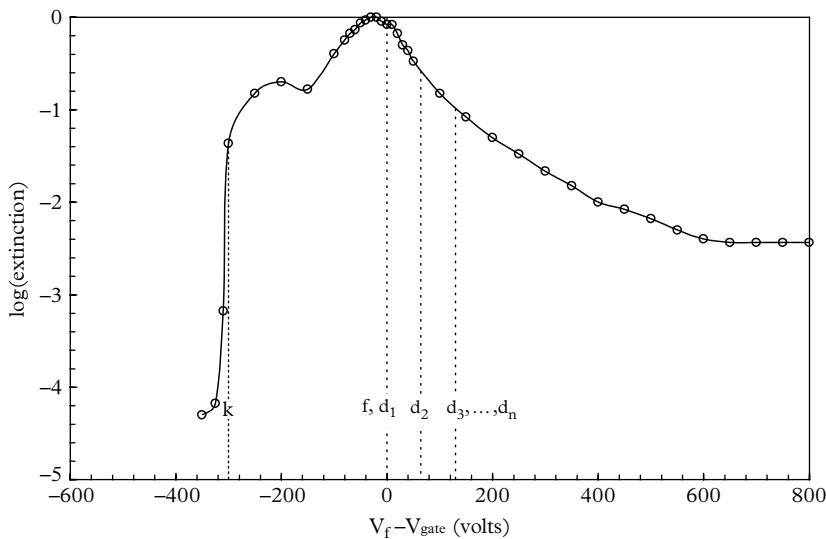


Fig. 13.44. Gating the focus electrode is particularly effective when V_f is set a few volts lower than the cathode potential.

this form of gating is not favoured. The effect of gating other dynodes is shown in Figs 13.42–13.44. A dynode close to the anode should not be gated because signal-electron densities are likely to be in saturation, resulting in afterglow and inter-dynode afterpulses.

13.9.5 Focus-electrode gating

Certain PMT types include a focus electrode disposed between the first dynode and the photocathode. The ET Enterprises 9214B series is an example of a fast PMT particularly suited to gating. Under normal operation, that is ungated, the innermost electrode in the front-end electro-optics is maintained at the potential of the first dynode, although not internally connected to it. Photoelectrons are prevented from reaching this dynode by switching the focus potential (pin 18) towards and slightly beyond the photocathode potential. This is illustrated in Fig. 13.44, which shows an extinction of 10^{-4} for settings beyond the photocathode potential. The transition is noticeably sharp, compared with dynode gating, since photoelectrons are emitted with only a few eV energy and are thus readily returned to the photocathode; in contrast, secondary electrons are significantly more energetic and not so easily captured. In the technical literature, the focus is often erroneously referred to as a grid or mesh. The ET 9124 series, for example, is specifically designed for gating the focus electrode, which is a hollow cylinder attached to a flange. Meshes are generally disfavoured in this application, for a number of reasons including reduction in collection efficiency and the generation of late pulses (see 11.3).

13.9.6 Circuit realization

The majority of gated applications reduce PMT gain, prior to and during an intense transitory light signal, thereby avoiding a long recovery from saturation. In an alternative mode, the PMT is normally on and switched off. It is clear from Fig. 13.44 that extinction of 10^{-4} is possible with the focus-switching technique alone, whereas comparable performance with dynode switching needs the combined contribution of at least two dynodes. Rise and fall times are important parameters that can be less than 10 ns, depending on the design. Where authors quote figures it is assumed, unless stated otherwise, that the standard 10 to 90% definition applies. However, Bristow *et al.* (1995) quote an elapse of 1 μs before the PMT returns to within $\pm 0.1\%$ of the quiescent gain; but, in terms of the standard rise time definition, switching is actually complete within tens of nanoseconds. Representative examples of gating configurations are given in Figs 13.45, 13.46, and 13.47. Extensive decoupling is necessary for all circuits but it is not shown for reasons of clarity—some authors decouple every stage.

13.9.7 Pulse generators and shielding

It is evident from Figs 13.45, 13.46, and 13.47 that pulse generators are, of necessity, AC coupled, thus limiting the permissible active time of the switch. This does not apply to the circuits discussed below, which are described by Hamilton (1971; Hamilton and Razi Naqvi 1973; Hamilton 2003) as, in these, only the fast trigger signals are AC coupled. The problem with most circuits is that the photocathode and the dynodes operate at high negative voltages, while the pulse generator is near the ground potential. The ET Enterprises GB1 series of gating products overcome this problem by deriving low-voltage power from the voltage divider, with the switching circuitry AC grounded and electro-optically isolated.

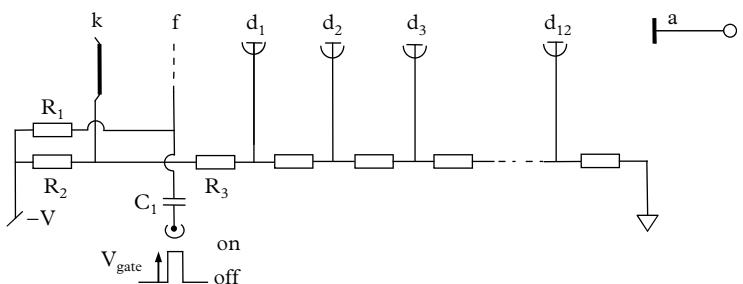


Fig. 13.45. To avoid droop in the gate signal, the time constant C_1R_1 must be long compared with the gate width. It is customary to set R_1 in excess of $1 \text{ M}\Omega$. The focus electrode, normally held at d_1 potential, is biased negatively with respect to the photocathode (determined by the ratio of $R_2/(R_3 + R_2)$). A positive gate pulse temporarily elevates the focus potential close to that of d_1 , thus enabling the PMT.

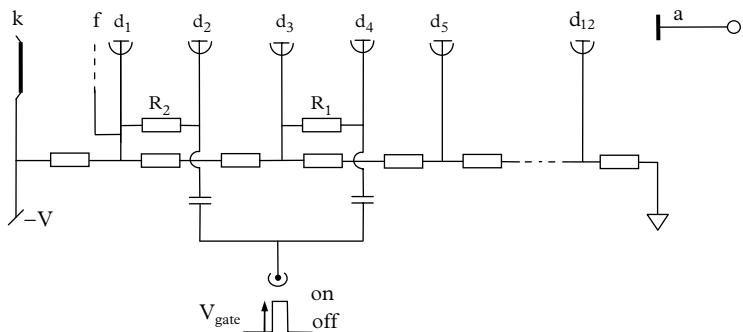


Fig. 13.46. An arrangement for switching d_2 and d_4 . Note that the gain of the PMT is sensitive to the amplitude of the gate pulse. Long-term stability can be achieved by satisfying the following relationship for d_2 : $V_{\text{gate}} = (V_{d2} - V_{d1})$ and similarly for d_4 : $V_{\text{gate}} = (V_{d4} - V_{d3})$. Similar reasoning applies to focus gating and, for the reasons given previously, R_1 and R_2 must be in the mega-ohm range.

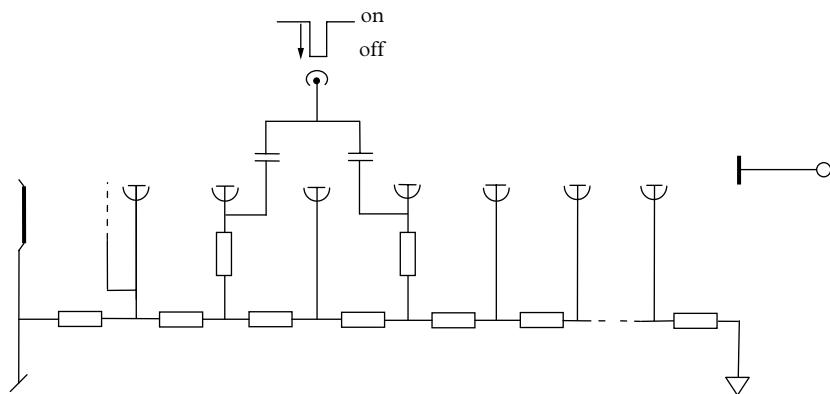


Fig. 13.47. An arrangement similar to that in the previous figure, but this circuit is normally 'on'.

Schulman (1989), Yoshida *et al.* (1989), and Piton *et al.* (1990), for example, also adopt this approach but that of Hamilton has three unique features. The first feature is that the dual outputs of a bistable are connected to adjacent dynodes; the PMT is active in one triggered state and, in the other, the dynodes are exchanged in polarity, generating an extinction coefficient of 10^{-3} . The second feature is the achievement of long gate times, determined only by the arrival of the next trigger pulse. The third feature is that the additional benefit derived from interchanging dynode potentials is the partial cancellation of switching transients observed at the anode.

Bristow *et al.* (1995) have reported on the deleterious effect of magnetic shielding in the form of a hollow cylinder. The operation of their gate was

Table 13.3 Summary of existing gating circuits.

Source	Mode	Gated electrodes	t_{on} (ns)	γ	Gate pulse
(1)	On→off	All even dynodes	80	10^{-4}	Blocking oscillator
(2)	Off→on	d_1	200	10^{-5}	Pulse generator
(3)	On→off	$d_2+d_3; d_7+d_8$	∞	10^{-6}	Bistable
(4)	Off→on	d_3+d_7	16	—	SCR Unitrode Corp.
(5)	Off→on	$d_1+d_2+d_6$	9	10^{-4}	Avalanche transistor 2N3019
(6)	On→off	All even dynodes	30	10^{-4}	HP214A pulse generator
(7)	On→off	f	50	10^{-4}	TMOS power MOSFETS
(8)	Off→on	d_1+d_4	50	10^{-3}	IRFF9220 switching FET
(9)	Off→on	$d_2+d_4+d_6; f; d_5+f$	1000	10^{-6}	MOSFET
(10)	Off→on	d_1	100	10^{-2}	TTL/optocoupler
(11)	Off→on	HV, cathode dynodes	100	10^{-4}	Thyatron

(1) Minami and Nishikawa (1966), (2) De Martini and Wacks (1967), (3) Hamilton (1971), (4) Ramirez and Kruse (1976), (5) Rossetto and Mauzerall (1972), (6) Wieme (1973), (7) Yoshida *et al.* (1989), (8) Piton *et al.* (1990), (9) Bristow *et al.* (1995), (10) Schulman (1989), (11) De Marco and Penco (1969).

unaffected when the connection to the shield was made directly to the photocathode. However, floating or the inclusion of a $10\text{ M}\Omega$ resistor, recommended by manufacturers as a safety feature, affects stability. Options available for shielding are discussed in 12.1.1 and illustrated in Fig. 12.4; the recommendation is one of finding the best configuration for shielding by testing the various options.

The electronic circuits listed in Table 13.3 for generating gating pulses are still relevant today, but with the possibility of improved performance. Power MOSFETs, for example, have steadily evolved regarding switching speed and voltage rating. Piton *et al.* (1990) incorporated these devices in the front end of a divider. A 500 V nanosecond-pulse generator using cascode-connected power MOSFETs is described by Baker and Johnson (1992). Optocouplers, a convenient way to isolate HVs, have also advanced in their switching speed.

REFERENCES

- Baker, R. J. and Johnson, B. P. (1992). A 500 V nanosecond pulse generator using cascode connected power MOSFETs. *Meas. Sci. Technol.*, 3, 775–7.

- Brigatti, A. *et al.* (2005). The photomultiplier tube testing facility for the Borexino experiment at LNGS. *Nucl. Instr. and Meth. in Phys. Res. A*, **537**, 521–36.
- Bristow, M. P., Bundy, D. H., and Wright, A. G. (1995). Signal linearity, gain stability, and gating in photomultipliers: Application to differential absorption lidars. *Appl. Opt.*, **34**, 4437–52.
- De Marco, F. and Penco, E. (1969). Pulsed photomultipliers. *Rev. Sci. Instr.*, **40**, 158–60.
- De Martini, F. and Wacks, K. P. (1967). Photomultiplier gate for stimulated-spontaneous light scattering discrimination. *Rev. Sci. Instr.*, **38**, 866–8.
- Duteil, P. *et al.* (1977). *Future Photomultiplier Assemblies and Associated Electronics in Large Experiments* (CERN-77-11). CERN, Meyrin.
- Flyckt, S.-O. and Marmonier, C. (2002). *Photomultiplier Tubes: Principles and Applications*. Photonis, Brive.
- Hamilton, T. D. S. (1971). Variable duration photomultiplier gating circuit. *J. Phys. E: Sci. Instr.*, **4**, 326–7.
- Hamilton, T. D. S. and Razi Naqvi, K. (1973). Instrument for time resolved phosphorimetry using an electronically gated photomultiplier. *Anal. Chem.*, **45**, 1581–4.
- Hamilton, T. D. S. (2003). *An Analogue Electronics Companion: Basic Circuit Design for Engineers and Scientists*. Cambridge University Press, Cambridge. http://www.et-enterprises.com/files/file/technical-information/rp061_gating%20of%20photomultipliers.pdf.
- Hiebert, R. D., Thiessen, H. A., and Obst, A. W. (1977). Photomultiplier tube base for high pulsed anode currents. *Nucl. Instr. and Meth.*, **142**, 467–9.
- Kerns, C. R. (1977). A high-rate phototube base. *IEEE Trans. NS*, **24**, 353–5.
- Kirillov, N. S. and Samokhvalov, I. V. (2014). ‘Application of an electro-optic shutter for strobing of lidar signals’. In *Proc. SPIE 92922D, 20th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*.
- Lagomarsino, V. and Testera, G. (1999). A gateless charge integrator for Borexino energy measurement. *Nucl. Instr. and Meth. in Phys. Res. A*, **430**, 435–46.
- Land, P. L. (1971). A discussion of the region of linear operation of photomultipliers. *Rev. Sci. Instr.*, **42**, 420–25.
- Lush, H. J. (1965). Photomultiplier linearity. *J. Sci. Instr.*, **42**, 597–602.
- Minami, S. and Nishikawa, K. (1966). A signal sampling photomultiplier circuit. *Appl. Opt.*, **5**, 173–4.
- Piétri, G. (1981). *Photomultiplicateurs*, Pub. RTC 5482–07–07/1981. R.T.C. La Radiotechnique-Compelec, Cedex.
- Philips. (1980). *Photo- and Electron Multipliers*. Phillips Photonics, Brie, France.
- Piton, M. C., Panning, W., and Winnik, M. A. (1990). An ‘on’-gated photomultiplier circuit for the determination of phosphorescence lifetimes. *Rev. Sci. Instr.*, **61**, 3726–28.
- Ramirez, J. J. and Kruse, L. W. (1976). Gated scintillator-photomultiplier neutron detector for time-of-flight measurements in the presence of a strong x-ray background. *Rev. Sci. Instr.*, **47**, 832–5.

- Rossetto, M. and Mauzerall, D. (1972). A simple nanosecond gate for side window photomultipliers and echoes in such photomultipliers. *Rev. Sci. Instr.*, **43**, 1244–6.
- Schulman, M. B. (1989). Gating circuit for linear-focused photomultiplier. *Rev. Sci. Instr.*, **60**, 1264–6.
- Takeuchi, S. and Nagai, T. (1985). Low power photomultiplier base circuit. *IEEE Trans NS*, **32**, 78–81.
- Wardle, R. (1982). *Gating of Photomultipliers*. ET Enterprises Reprint Series, RP061. ET Enterprises Ltd, Uxbridge. Available at http://www.et-enterprises.com/files/file/technical-information/rp061_gating%20of%20photomultipliers.pdf (accessed 6 Feb 2017).
- Wieme, W. (1973). A versatile circuit for light intensity measurements with a gated photomultiplier. *J. Phys. E: Sci. Instr.*, **6**, 203–5.
- Yoshida, T. M., Jovin, T. M., and Barisas, B. G. (1989). A high speed photomultiplier gating circuit for luminescence measurements. *Rev. Sci. Instr.*, **60**, 2924–8.

14

Electronics for PMTs

14.1 Introduction

PMTs are high-gain, wide-bandwidth transducers capable of detecting single photons. No other device offers the combination of a large detection area, gain of up to 10^9 , and sub-nanosecond response. The fastest OP-AMPS currently available are characterized by gain-bandwidth products that only approach those of PMTs. It is therefore natural to query why additional gain should be necessary with a device of the stated attributes. The answer is that there are signal transformations beside gain when PMTs form part of a detection system, the optimal functioning of which may depend on smoothing and shaping the signal using amplifiers with these properties. For example, nuclear spectroscopy systems based on scintillators are invariably modular, consisting of a crystal scintillator, a PMT, a preamplifier, a shaping amplifier, an MCA, and a display. Shaping amplifiers are also referred to as filter, or spectroscopy amplifiers, but essentially they all perform the same function, that is, they provide some degree of signal conditioning in addition to selectable amplification. The aim of the user should be to use appropriate electronic modules, whatever detector system is to be realized, and then to ensure that each module performs within its optimal range of operation. It makes no sense to operate an electronics amplifier ‘flat out’ with the PMT ‘coasting’, and vice versa, since better performance will always result from an equitable gain allocation. A difficult challenge may lie in having to use non-optimal ‘existing laboratory electronics’, in which case the relative sensitivities of the constituent modules is even more critical. PMTs conform closely to an ideal current generator, and an impedance transformation to a voltage generator may be achieved by adding a load resistor in series with the anode; a $50\ \Omega$ resistor is the obvious choice for matching standard coaxial cable—a purely passive solution. Alternatively, current to voltage conversion may be achieved with a transimpedance amplifier, ideally mounted in close proximity to the PMT. Such an amplifier presents a short circuit to an anode signal, and provides an output voltage from a low-impedance source. There is a charge-sensitive variant of the transimpedance amplifier that converts charge to voltage, discussed in 14.9.1.

Traditionally, measurement of low-level DC and slowly varying signals has been the preserve of expensive and delicate electrometers connected directly to the anode. These instruments have a history of over 100 years, following the invention of the vacuum triode in 1906. Early models were known as vacuum tube voltmeters, but the products available nowadays utilize JFETs or MOSFETs to provide very high-input impedance devices of up to $10^{16}\ \Omega$. This means that they can pick off a voltage at the picovolt level without loading the circuit, and currents as low as 10^{-17} amps can be measured. Instruments with such high sensitivity lie at the top end of instrumentation: handling them requires skill, know-how, perseverance, and, ideally, years of experience. Fortunately, non-veterans may turn to an excellent publication on the subject produced by Keithley Instruments Inc.: *Low Level*

Measurements Handbook (7th ed.) (Keithley Instruments Inc.1972). This is essential reading for anyone planning to make ultra-low-level DC measurements.

The widespread availability of inexpensive, battery-operated 3½-digit multimeters provides a convenient means for making current measurements. In practice, this is achieved using the voltage range and an anode load of the order of 1 MΩ. Furthermore, because these instruments can be floated, the user has the choice of adopting either HV polarity. However, care has to be taken with positive HV to avoid an obvious electrical shock hazard. Most multimeters of the type described here can also measure true rms voltages, a facility the author has found useful for noise and linearity investigations.

An extensive historical literature on the subject of electronics for PMTs exists, much of which is concentrated in a few scientific journals: *Nuclear Instruments and Methods*, *Institute of Electrical and Electronic Engineers (IEEE) Nuclear Science, Applied Optics, Journal of Physics E*, and *Review of Scientific Instruments*. These journals cover PMT applications in general, and a reader new to the subject is advised to consult such publications before embarking on any electronics design: the likelihood is that ‘your particular electronics problem’ has already been solved. The number of textbooks on the subject is limited but the author has found those by Knoll (1989), Nicholson (1974), and Herbst (1970) to be particularly useful. A set of lecture notes (available from the web) on radiation detectors and signal processing by Spieler (2001) gives extensive coverage of detectors and associated electronics. NIM catalogues are available on the web from manufacturers such as EG&G ORTEC (now part of AMETEK), Canberra, and CAEN. These excellent sources of information on nuclear electronics are available as downloadable library items, although original hard copies are highly desirable. A salutary exercise, recommended at an early stage of detector development, is to capture PMT signals on a digital oscilloscope. The need, or otherwise, for some degree of signal conditioning becomes immediately apparent; particularly for pulsed signals containing a small number of photons, such as those produced by low-energy X-rays captured in NaI(Tl), for example. Photoelectron numbers are even lower and the spread in arrival times and their variable amplitudes, after multiplication, give rise to ragged output signals. The information relevant for spectroscopy purposes is the number of photoelectron-initiated pulses contained in each scintillation event. Given a sufficiently fast PMT and a digital oscilloscope, it is just feasible to count this number, to provide the best possible measure of the size of each event. However, signals that overlap in time are not resolved as a pair of events but as a single spurious one.

DSP is a firmly established technique for use with PMTs; all the traditional circuit concepts developed, by what we can now refer to as ‘classical’ electronics, can be achieved by DSP—but faster and better. The approach adopted in this chapter to signal manipulation is based mainly on the classical development of the subject, from which the underlying digital approach derives. In DSP the PMT output is digitized continuously, analysed on-line to extract the relevant information, and then discarded as new information is presented. Contributions to dead

time in traditional photon-counting systems originate from pulse overlap and from discriminator recovery, (4.19). In DSP systems, there is no contribution from discriminator recovery time because algorithms have been developed for separating overlapping anode signals. There is a considerable literature on this subject, with descriptions suitable for the non-specialist by Pasquali *et al.* (2007); they describe a DSP system devised for particle identification in CsI(Tl) scintillator. A further account of DSP is given in 14.14, based on the availability of excellent manufacturers' application notes.

14.2 Charge output transducers

There is a diverse range of detectors, including PMTs, that produce charge in response to an input stimulus. In certain applications the only measured parameter of concern is the total charge generated by each event, but in others the time signature of individual events may be required. A preamplifier can be designed to produce an output voltage, V_o , proportional to Q_a , where Q_a is the total charge generated at the anode. Alternatively, referring to Fig. 14.1, the preamp can be devised such that $v_o(t)$ has the same time signature as $i_a(t)$. In this case, $i_a(t)$ is scaled by R over the duration of the event, where $i_a(t)$ represents the generation of charge with time. Note that $v_o(t)$ embodies the profile of $i_a(t)$ and, if the anode of a PMT is connected to a pure resistance, $v_o(t) = R i_a(t)$. The latter operation is referred to as a fast or current preamplification. Transducers may be classified as either direct charge generators, or those that include an intermediate process of light emission. All detectors listed in Table 14.1 are used in conjunction with a preamplifier to convert output charge to a voltage analogue. With the exception of Geiger–Müller (G–M) tubes and APDs, operated in the avalanche mode, these detectors are linear, or at least linear over a useful range of input excitation.

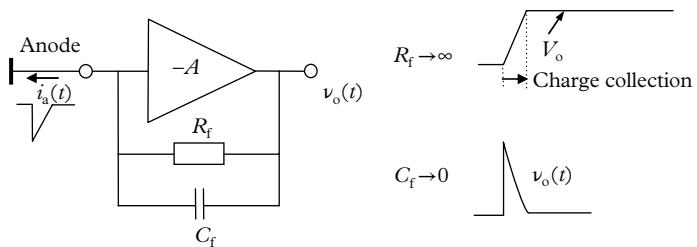


Fig. 14.1. The generic description of transimpedance preamplifiers; A is the open loop gain.

Table 14.1 Direct charge output detectors.

DC output detectors	Light-to-charge converters
G–M tubes	PMTs
Ionization chambers	APDs
Large-area wire chambers	SiP arrays
Lithium-drifted silicon detectors	PIN diodes
Germanium detectors	
Si(Li) detectors	
Cadmium zinc telluride (CZT)	
Channeltrons	

14.3 The role of a preamplifier

The distinction in the technical literature between preamplifiers and amplifiers is not always clear. In the context of PMT applications, a preamplifier serves as an interface between a PMT and a shaping amplifier, or some other electronics unit, such as a discriminator or an SCA. Its primary role is that of a transimpedance device, converting a high-impedance current source, the PMT, into a low-impedance voltage generator. This is necessary because electronic circuits, whether discrete or integrated, operate primarily on voltage signals. The conversion of current to voltage can be made with a single resistor connected between anode and ground, but the signal is then sensitive to loading, should additional circuitry be added. An essential requirement of a preamplifier is that its output signal is suitable for subsequent amplitude or time measurement, noting that the magnitude of the total charge and the time profile are independent sources of signal information. Reference to the dimensionless parameter, gain, is inadmissible with this type of amplifier because the units for the input and the output are dissimilar. The terms conversion gain and transimpedance are generally preferred for their description. Further to the considerations in 14.2, the nature of the output is determined by whichever of the two circuit elements dominates performance: if R_f is sufficiently large, then the capacitor determines the sensitivity, yielding a peak voltage $V_o = -Q_a/C_f$ that slowly decays back to the baseline. The output signal profile of the preamplifier is essentially independent of the anode signal time structure: all output pulses have the same shape and decay time constant. On the other hand, when C_f is relatively small or absent, R_f determines the performance and, in the extreme case, $C_f \rightarrow 0$; $v_o(t) = R_f i_a(t)$; and the shape of the input current signal is replicated as an output-voltage analogue.

In commercial charge-sensitive preamplifiers, C_f ranges from 0.1 to 5 pF, and R_f is chosen within the range 100 MΩ to 1 GΩ; the resulting time constant is 1 ms. Note for typical values of $R_f = 1$ GΩ, and $C_f = 1$ pF, the resulting discharge time constant is 1 μs. For the intended applications, this time constant is many orders of magnitude greater than the charge collection time, which is about 1 μs for NaI(Tl), for example. A high-pass filter is therefore included to reduce the time constant to be compatible with that of the signal. The role of R_f is to provide DC stability and to discharge C_f between events. It will be shown that R_f should be of high value to attain the lowest noise, and some other means has to be found to reduce the time constant of the signal discharge tail; this is accomplished by imposing a faster decay time in subsequent circuitry (see 14.9.2, Fig. 14.14). Closed loop gain, a dimensionless parameter in circuit analysis, has no meaning in this configuration because the preamplifier operation is one of charge-to-voltage conversion; an operational sensitivity is the relevant parameter, in units of volts per picocoulombs when C_f dominates and in volts/mA when R_f controls the feedback.

In fast preamplifiers, R_f lies within the range of 100 Ω to 50 kΩ, with C_f selected between 0 and 5 pF. The sensitivity of a preamplifier, that is, its ability to produce the maximum conversion gain, varies as $1/C_f$ and it is common to come across capacitance values as low as 0.2 pF; in addition, a discharge resistor, R_f , is placed in parallel with C_f . These amplifiers are primarily intended for use in systems where 50 Ω input and output matching applies and hence it is appropriate to quote gain in volts per volt. A gain of 10 is typical for a 200 MHz bandwidth preamplifier, achieved by selecting $R_f = 500$ Ω. Intermediate operation, where both R_f and C_f are effective, is an option where speed (rise time) may be sacrificed to attain a degree of waveform smoothing, characterized and quantified by a time constant, $\tau_f = R_f C_f$. The use of a transimpedance preamplifier developed by Goyot (2000) for APD pulses of ~20 ns width is an example of intermediate operation for which a τ_f of 8 ns applies. The transimpedance amplifier used by the MAGIC collaboration has a 2000 Ω feedback resistor with C_f omitted. There is always an associated fraction of a picofarad stray capacitance, and this amplifier produces pulses of 6 ns width. These topics are covered in detail in 14.11.

14.3.1 Interfacing and signal transmission

It is appropriate at this juncture to discuss the general topic of interfacing a PMT to electronic circuitry. The discussion that follows centres around connecting a preamplifier to a PMT: either adding it to the voltage divider board or retaining it as a standalone unit connected to a PMT by a coaxial cable. The latter option is adopted in commercial units, such as the NIM range of modules, or for a customer's own circuitry. This applies to the use of other types of electronics, such as discriminators. Interconnecting standalone electronics units of whatever type is usually carried out using coaxial cable; although twisted pair may be used for less demanding applications (it has a characteristic impedance of ~100 Ω). Undesirable signal reflections occur in those situations where matching is

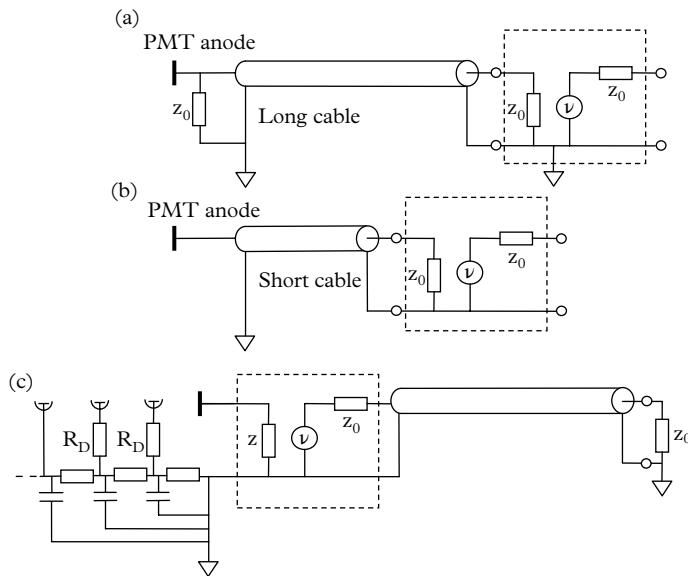


Fig. 14.2. Connecting an amplifier to a PMT (enclosed in the dashed box): (a) using a coaxial lead of length greater than 20 cm; (b) using a short cable of less than 20 cm; (c) mounted close to the anode of the PMT. The same considerations apply to other types of electronic circuits, such as discriminators and ADCs.

unfulfilled. Coaxial cables are available with a characteristic impedance z_0 ranging from 50 to 93 Ω . In practice, it is customary to terminate both ends of a cable with a matching resistor, z_0 , as in Fig. 14.2(a). In Fig. 14.2 (b), the matching resistor at the anode has been omitted, arguing that the far end is terminated in any case. The advantage of using a short cable with a single termination is that the received signal is twice the amplitude of the fully terminated one. In theory, there should be no signal reflections from a matched termination, but perfection is difficult to achieve at any termination because of the presence of low-amplitude reflections. Figure 14.2(c) illustrates the advantages of locating a high-impedance preamplifier close to the anode, as is recommended for charge-sensitive and high-impedance preamplifiers. Despite matching, up to 10% of the received signal is returned to the anode, which, if open circuit, will then transmit the full reflected component to the other end of the line. Propagation time is typically 5 ns/m and, if a cable is less than 20 cm in length, the reflected contribution is generally lost in the tail of the main pulse. A mismatched cable greater than 1 m in length will produce a satellite pulse separate from the main pulse and which may have sufficient amplitude to be counted or encoded. Furthermore, the reflected pulse may be of either polarity, depending on where the mismatch lies.

Figure 14.2(c) illustrates an amplifier mounted in proximity to the anode, without the need for coaxial cable. With a high-input impedance voltage amplifier

(e.g. see 14.11.1, Fig. 14.22), it is possible to operate with $z > z_0$, leading to a corresponding increase in system gain. The circuit in Fig. 14.2(c) is configured for negative HV applied to the photocathode. Resistors marked R_D are included for damping purposes (discussed in 13.8.8). The sudden discontinuity created at both ends of a coaxial cable (lack of dielectric) compromises matching, particularly where a low-cost $50\ \Omega$ terminator, such as one consisting of a radial resistor enclosed in a metal barrel, for example, is used. Improved matching can be achieved by including a terminating pad consisting of a combination of small-value inductors and capacitors (0.1 to $1\ \mu\text{H}$, and 1 to $10\ \text{pF}$). This is largely a forgotten art but, fortunately, the classic amplifier papers discussed in 14.11.1 include examples of appropriate trimming circuits.

14.3.2 Grounding and shielding

Low-light-level applications, particularly those involving charge-sensitive pre-amplifiers, are particularly susceptible to noise pick-up, leading to unstable performance—they demand special attention to grounding. Massive detectors, found in astrophysics and accelerator experiments, generate additional problems exacerbated by long cabling. Unfortunately, this vital topic is seldom reported upon in the scientific literature but is usually relegated to that of an internal file note. Some reports, such as *DIRC Grounding and Shielding* (DIRC 1997), are recommended reading.

Ground-plane construction is mandatory for fast electronics. The key consideration is a star ground connection for all of the following: the anode signal coax braid; the HV outer braid; and decoupling capacitors, as illustrated in Fig. 14.2(c). PMTs are invariably mounted within a light-tight metal enclosure. A cylindrical construction is preferred for a single PMT where the anode signal, HV, and any low voltage connections enter through an end cap; in addition, there may be a requirement for a dynode signal. This compact arrangement appears to work well. However, deploying a plurality of PMTs within a relatively large metal enclosure may generate ground loops. These can be avoided by using insulated feedthrough bulkhead connectors of the BNC type for signal transmission. Ott's (1988) book is one of a few to offer guidance where crosstalk and noise pose a seemingly insurmountable challenge.

14.4 Circuit protection and limiting

HV transients are generated whenever the HV is switched on or off (see 13.7.4). Time constants are of the order of milliseconds for either HV polarity but, because of the proximity to the anode, positive HV is more hazardous than negative

polarity (note, R_L is connected directly to the anode). Delicate electronics are easily damaged should the HV supply suffer catastrophic failure. The PMT may experience intermittent breakdown, because of excess voltage. Other causes of breakdown are high-light-level overload; operation outside the specification; and loss of vacuum within the PMT. Perhaps the most destructive transient is that generated when an electronics unit is inserted or removed from an active PMT. When disconnected from the electronics, or an oscilloscope probe, the anode, if floating, charges to either the potential of the last dynode or to positive HV, depending on the polarity of the supply. On reconnection, these potentials are discharged through the electronics, and sometimes with destructive consequences.

Protective measures can be taken to remove the insertion risk. A $10\text{ M}\Omega$ resistor permanently connected between anode and ground, where negative HV is used, ensures that the output is always at ground potential; for positive HV, a safety resistor is connected from the output side of the coupling capacitor to ground. In most cases, a $10\text{ M}\Omega$ resistor is imperceptible to the electronics. However, there are some low-light-level applications concerning electrometers in which the inclusion of a resistor of any value is unacceptable, in which case, safeguarding the electronics is a matter of self-discipline. Other measures based on signal diode protection are usually incorporated within commercial instrumentation, examples of which are given in Fig. 14.3. The action of the diodes is to divert excess current to ground, whether positive or negative. The input voltage is restricted to $\pm 0.7\text{ V}$, which is representative of the forward voltage drop of a silicon signal diode. This basic circuit offers a limited safeguard, as the diodes themselves need protection from any higher-amplitude transients. More advanced protection circuits are shown in Fig. 14.3(b) and (c); these use matched diodes, which are biased to draw a few millamps continuously. Protection extends to signals approaching the reverse breakdown voltage of the diodes, typically 100 to 200 volts. Circuit (d) is an example of a soft limiter that prevents the output of an OP-AMP from entering saturation but it does not offer protection against transients at the anode in the manner of the other examples. The onset of limiting is determined by the ratio of the two marked resistors (R and $3R$) and on the supply voltage; limiting occurs at about 2 V for a 6 V supply in the circuit shown.

14.5 Modular electronics

Every light detector needs the appropriate electronics. An outline of a generalized system is given in Fig. 14.4. In certain applications, the electronics may comprise just an oscilloscope, connected directly to the anode; nowadays, this is likely to be a digital instrument with computational capability. In the disciplines of high-energy

physics and astrophysics, the scale of the detectors, combined with the demands required of the electronics, has led scientists to design their own electronics, usually in the form of a hybrid circuit or an application-specific integrated circuit (ASIC). An alternative to self-construction is to take advantage of one of the by-products of the nuclear industry. A suite of NIM units is available from several manufacturers, enabling the user to configure measurement systems for a range of applications. These modules conform to a standard, making it possible to incorporate products from different manufacturers within the same system. Units slot into a bin 19" wide that can accommodate and provide power for up to 12 standard width modules; mini-bins, 12" in width, are also available. The primary benefit of NIM units, for someone working in a facility where they have already been adopted, is the speed and convenience with which a system can be assembled or modified. The drawback in such an environment is that modules that are fortuitously available may be used inappropriately. Modular systems can also be

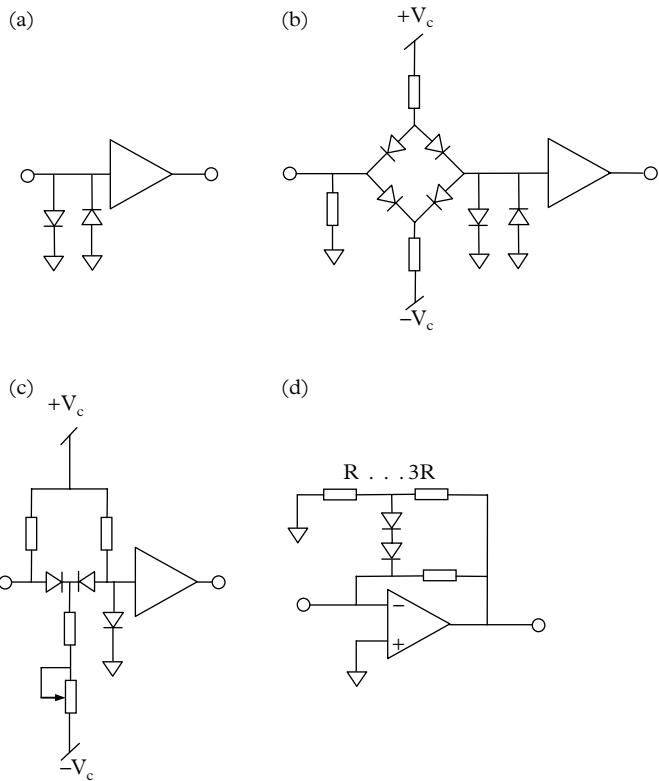


Fig. 14.3. (a), (b), and (c) illustrate some of the ways in which to protect electronic circuits from HV transients; (d) is an example of a limiter that hastens recovery from signal overload.

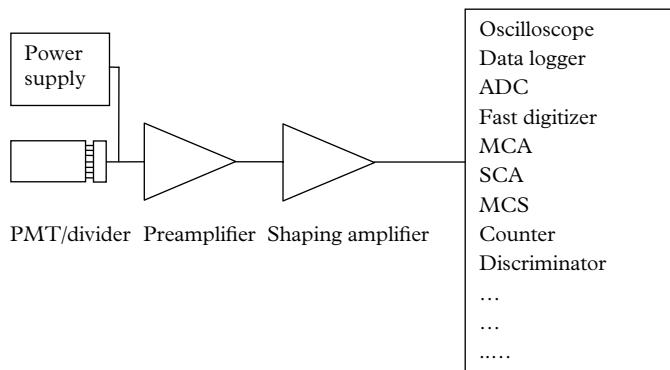


Fig. 14.4. Interconnected electronic modules constitute a complete measuring system. A selection of available output devices is shown in the box.

constructed using CAMAC (computer-automated measurement and control) units. These products are intended for large-scale digital systems, usually involving ADCs and bussing of data under computer control. More detailed information on the two instrument standards may be found in Knoll (1989), or from manufacturers. CAMAC and NIM units are available covering the following major instrumentation categories:

- preamplifiers[†]
- amplifiers
- fast-timing discriminators
- TACs
- MCAs and multichannel scalers (MCSs)
- counters
- timers
- ratemeters
- SCAs
- delays
- pulse generators (fast step function)
- HV power supplies

Electronics need not be complicated: the degree of sophistication depends on the demands of the application. A typical example might include a light detector, a preamplifier, a shaping amplifier, and some type of peripheral device to handle output from the electronics. Recall from 14.3 that a preamplifier's primary function is to condition the signal received from a PMT so that it is suitable for the next

[†] For high-sensitivity applications, preamplifiers are best mounted close to the PMT and, for this reason, some are packaged in a small metal box and powered remotely from an NIM bin or from a shaping amplifier.

module of the system. This module may be an amplifier or a shaping amplifier, the first of which performs only amplification, while the second includes filtering and pulse shaping as well. Charge-sensitive preamplifiers are ideally mounted on a circuit board closely coupled to the PMT output in a way that minimizes stray capacitance and avoids long signal-transmission cables. In less demanding applications, the preamplifier may be remote from the PMT, with its input connected by a short length of coaxial cable. There is always a small loss in gain with remote coupling, but the real limitation on the use of long unmatched cable relates to capacitance. This is of the order of 50 to 100 pF/m, and the principal concern is one of noise generation. The demands placed on the performance of electronic circuitry are usually minimal where detected light signals are of restricted bandwidth, relatively intense, and of low rate, but dealing with fast low-level-signals calls for specialized electronics. PMT applications fall into two broad categories: in 14.9 we are concerned with measuring the total charge content in each signal, while in 14.11.1 to 14.11.5, the primary concern is that of fast timing. It is feasible to perform both of these functions with a single PMT by either splitting the signal before or just after the preamplifier or by taking two outputs from a single PMT (see section 14.7.)

14.6 PMT equivalent circuit

Understanding the equivalent circuit of a PMT is a prerequisite to making the right choice of ancillary circuits. PMTs conform closely to an ideal current generator with parallel resistance R_0 and self-capacitance C_0 ; typically, R_0 exceeds $10^{10} \Omega$, and C_0 is generally less than 20 pF. The resistance derives from traces of alkali metals on surfaces and pin connections within the PMT envelope. Application of biasing voltages generates leakage currents, and those that terminate on the anode electrode, or its pin connection, are particularly undesirable (see 6.3). PMTs incorporating anode guard rings have been offered in the past, primarily as development devices, but without sustained commercial production: the implication being that electrical leakage in PMTs is generally not a primary limiting factor. One of the essential PMT attributes is an ability to distance a signal from this leakage component, based on high PMT gain: leakage currents are not amplified by the multiplier, and their magnitude, at the anode, varies roughly in proportion to the HV, V . On the other hand, the signal amplitude varies as V^β , with $\beta \approx 8$, and hence the signal-to-leakage ratio will vary as $V^{\beta-1}$, allowing the signal to exceed the leakage current at sufficiently high gain. Many users fail to appreciate this fundamental point. Vacuum photodiodes are unity gain devices that do not have the property of multiplier gain, making them essentially high-light-level detectors. However, they conform to the same PMT equivalent circuit and therefore it will be obvious that some of the considerations already covered are also relevant to vacuum photodiodes.

A current generator producing a transient signal is assumed in the following analysis. The same equivalent circuit applies for continuous signals, except that the current generator is now excited by a DC signal. The circuit analysis that follows is based on the response to a transient input signal of exponential form. The assumed signal profile is idealized since all PMTs have finite rise time, but this detail can be safely ignored for most analysis purposes. Where fast timing is concerned, it may be necessary to adopt a more representative form of the input stimulus, such as $f(t) = t_n \exp(-t/\tau_i)$. This function is an n -fold convolution of an exponential and it closely mimics the physical multiplication process in an n -stage PMT (Shubnikov and Subbotin 1972). The output signal, $f(t)$, is characterized by a finite rise time with a slower fall time, analogous to the behaviour of actual PMT pulses. However, circuit analysis is more complicated for this form and is not covered here. The important attribute of an exponential function, from a purely practical standpoint, is that it is easily integrated and differentiated and yet is sufficiently close to actual PMT behaviour to justify its adoption. The concern for the moment is pulse shaping, noise, and pulse height encoding, for which the time constants associated with the electronics are long compared with typical PMT rise times, and the adoption of an exponential function is certainly valid in this case. The appropriate time constant, τ_i , for investigating single-photon events can be taken as t_{fwhm} , as given by PMT manufacturers (t_{fwhm} ranges from 100 ps to about 30 ns, depending on the PMT type). We can establish the magnitude of single-photoelectron initiated events by assuming a multiplier gain of 10^7 —a typical setting for this application. The mean charge delivered to the anode, Q_a , is 1.6 pC by each photoelectron. When investigating the response to a scintillator flash, from NaI(Tl) for example, Q_a may be taken as 10 pe/keV of absorbed X-ray or gamma radiation, with $\tau_i = 240$ ns. Most scintillators can be characterized in this way by choosing the appropriate pair of parameters for amplitude and time. The PMT gain required, when used with industry standard preamplifiers and shaping amplifiers, turns out to be in the region of $\sim 10^5$ for NaI(Tl), $\sim 10^6$ for plastic scintillators, $\sim 3 \times 10^6$ for Cerenkov converters, and $\sim 10^7$ for single-photon events. The input excitation for all circuitry coupled to the anode is taken as

$$i_a(t) = -\frac{Q_a}{\tau_i} \exp(-t/\tau_i), \quad (14.1)$$

which is normal to $\int_0^\infty i_a(t) dt = -Q_a$, the total charge in the anode pulse. Note

that (14.1) is quantitatively correct so that its adoption will account for both amplitude and time structure. For single-photon excitation, τ_i ranges from ~ 100 ps for MCP detectors to ~ 30 ns for box-and-grid dynode structures. Scintillators can be characterized in the same way, with τ_i ranging from ~ 600 ps (BaF_2) to ~ 1 μs for CsI(Tl). Cerenkov light emission is very fast and, for small-volume detectors, the appropriate value for τ_i can be assumed to be that of the PMT. For massive detectors, τ_i is determined by a combination of path length and PMT contributions. It is perhaps worth mentioning here that fast response in a PMT is

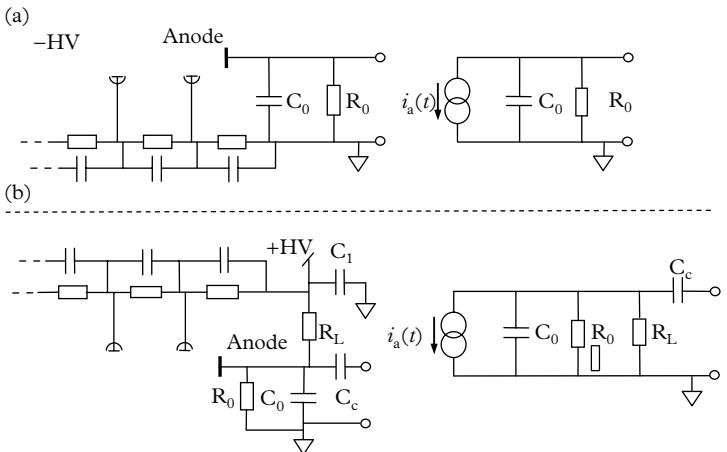


Fig. 14.5. (a) The equivalent circuit for a PMT operated with negative HV, showing the connections to an external circuit; (b) refers to +HV operation. Capacitor C_1 must have sufficient capacitance to ensure that the HV is at AC ground, but it plays no part in circuit analysis.

not always an advantage: the electronics for use with a slow PMT is considerably easier to develop and it may be a mistake to choose a fast PMT only to find that the existing electronics is incapable of handling faster signals.

The anode is always maintained at a positive potential with respect to all other PMT electrodes. This leads to two options for biasing: either the anode is operated at ground potential, in which case the cathode assumes a high negative potential, or the cathode is grounded, with the anode held at positive potential. Consequently, there are two equivalent circuits shown in Fig. 14.5. Considerations that favour the use of one polarity rather than the other will now be discussed.

When an application involves measurement of continuous, as opposed to pulsed, signals, the mode of operation places the anode at ground potential. Where signals consist of pulses, usually of short duration, the user has the option of operating with either positive or negative HV. The favoured choice in high-energy physics applications is negative HV with the coupling capacitor, C_c , omitted; this provides direct access to the anode. This choice of polarity offers two advantages over the alternative. First, the base line remains fixed at 0 V, free from the effects of variable event rates—a phenomenon examined in some detail in 14.8.2 and 14.9.3. Second, a capacitor attenuates the low-frequency response of the electronic system, thereby affecting the information in a transient waveform. Distortion-free response is important in PSD, LIDAR, and fluorescence studies, for example.

Applications involving positive HV require a capacitor, C_1 , to prevent HV ripple from reaching the load resistor R_L , shown in Fig. 14.5(b). This resistor performs the conversion from current to voltage. Since pulsed and DC applications can be satisfied with negative HV, the benefits of the positive HV option may

not be obvious, but the explanation lies in PMT history and in the way the devices actually perform in practice. All manufacturers warn that anything, especially metal constructions, in close proximity to the window of a PMT must be held at cathode potential, and the penalty for disregarding this is unstable performance. An example for which the industry standard is positive HV is found in scintillation counting using NaI(Tl) crystals. These deliquescent scintillators are hermetically sealed within a metal can; operation with the enclosure at other than cathode potential portends practical difficulties, not the least of which are safety related.

14.6.1 PMT signal bandwidth

Any discussion concerning fast pulse amplifiers is likely to involve the signal bandwidth. For a DC-coupled amplifier, with a single high-frequency pole at f_h , the time constant is $\tau = RC$, and $f_h = 1/(2\pi\tau) = B$. The bandwidth is also given by $B = 0.35/t_r$, where t_r is the rise time of the output for a step input. This is explained in detail by Millman and Halkias (1972). The rise time of the output, for a step of magnitude v_o , is dictated by τ in accordance with

$$v_o(t) = v_o(1 - \exp(-t/\tau)).$$

The time taken for the signal to reach nine-tenths of the final amplitude, minus the time taken to reach one-tenth, is defined as the rise time, t_r , of $v_o(t)$:

$$t_r = \tau \ln(0.9/0.1) = 2.2\tau,$$

and so

$$B = 1/(2\pi\tau) = 2.2/(2\pi t_r) = 0.35/t_r.$$

Consequently,

$$t_r = 2.2\tau = 2.2/(2\pi\tau) = 0.35/B. \quad (14.2)$$

The rise time for a bandwidth of 100 MHz is therefore 3.5 ns. It is important, however, to remember that the relationship between B and t_r refers to excitation by a pulse with an exponential profile of time constant τ . Application to real PMT pulses, with finite rise and fall times, should be regarded as providing only an order of magnitude indication of B . The frequency response of a PMT may be determined experimentally by analysing the anode output initiated by a single-photoelectron event, as presented in Fig. 14.6: that is, a delta function input, $\delta(t)$. The impulse response $h(t)$ in Fig. 14.6 leads to the frequency response, via the Fourier transform of $h(t)$. The same procedure applied to a fast multi-photoelectron Cerenkov pulse produces a broader output pulse because of the variation in the $k-d_1$ transit time of individual photoelectrons. This arises when the entire window of the PMT is illuminated. Only a single shot of $h(t)$ is necessary to uncover the frequency responses of the PMT, as shown in Fig. 14.7. Note, for single-photon excitation, the $k-d_1$ transit-time dispersion has no effect because every event is initiated by a single electron launched from d_1 .

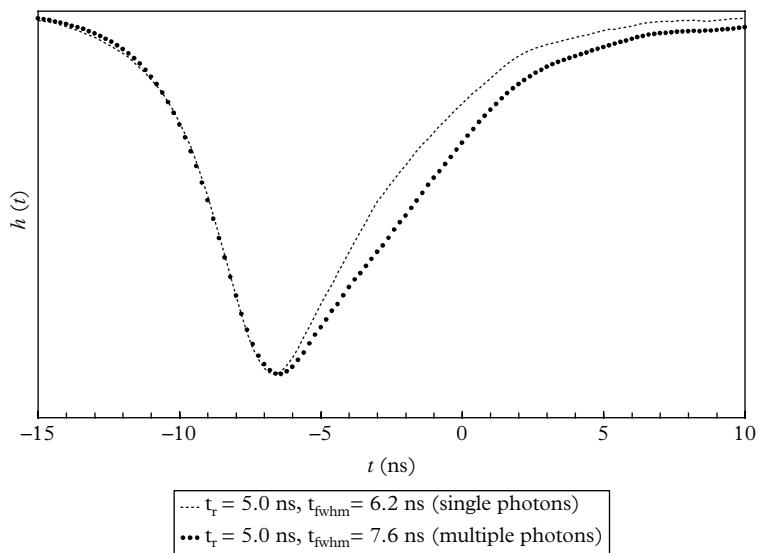


Fig. 14.6. Single- and multi-photoelectron signals for a medium-speed PMT, Type 9266B. The leading-edge time response is common to both curves but multi-photoelectron pulses have an extended tail.

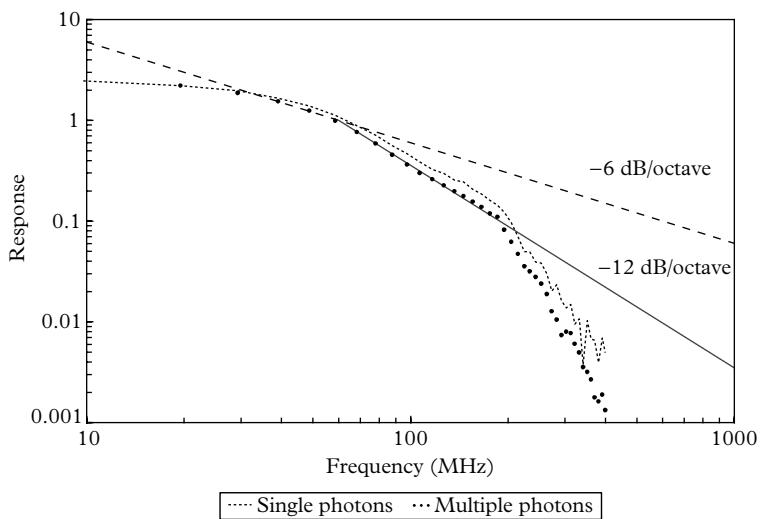


Fig. 14.7. Bode plot for a 9266B, medium-speed PMT, derived from the Fourier transform of Fig. 14.6. The bandwidth, B , is about 40 MHz at 3 dB down, compared with 70 MHz, based on the approximate formula $B = 0.35/t_r$.

14.7 Multiple PMT outputs

There are applications in which both energy encoding and precise timing are simultaneously required. This may be done in the way shown in Fig. 14.8(a) by picking off the anode signal with preamplifier A_1 , followed by signal conditioning in A_2 to suit energy measurement, and similarly in A_3 for timing purposes. The primary requirement is that the input impedance of A_2 and A_3 should be significantly higher than the output impedance of the preamplifier, thus minimizing loading and crosstalk; A_1 is either a charge-sensitive preamplifier or a voltage amplifier, whereas the choice of A_2 and A_3 is restricted to voltage amplifiers. Note that, with DSP, any number of copies of the output of A_1 can be generated for subsequent manipulation.

A useful attribute of a PMT is the availability of multiple signal outputs. In addition to the output from an anode signal, outputs may be taken from one or more of the dynodes. These signals are positive-going and of reduced amplitude, but otherwise they are a faithful reproduction of the anode signal; note that, at this late stage of the signal forming process, statistical effects cease to be important because the final shape of the signal is determined by earlier stages. The charge, or pulse height, of a signal taken from the last dynode of an n -stage PMT has a relative amplitude of $(\delta_n - 1)/\delta_n$, which is $\sim 80\%$ of the anode signal; the output from the penultimate dynode is also positive-going and has a relative amplitude of $(\delta_n - 1)/(\delta_n \delta_{n-1})$, which is $\sim 20\%$ of the anode signal amplitude. Dynode signals are developed across a load resistor, R_L , as shown in Fig. 14.8(b), which illustrates

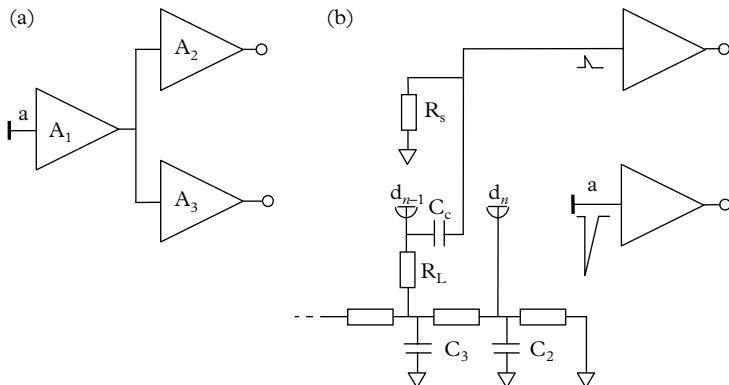


Fig. 14.8. Configurations for deriving two simultaneous signals from a PMT. The role of A_1 is to convert a current generator (the PMT) to a voltage generator, in addition to providing a low-impedance voltage source for A_2 and A_3 ; A_2 may be configured for spectroscopy purposes, while A_3 conditions signals for timing purposes. The equivalent circuit for dynode pick-off is that in Fig. 14.8(b).

pick-off at the penultimate dynode. The use of this dynode is preferred over using the ultimate one because the former offers lower crosstalk and better isolation from the anode signal, and this is especially relevant under overload. Also, stray capacitance at d_{n-1} is about half that between d_n and the anode; such a reduction in capacitance has a significant effect on noise (see 14.10.2) and on signal matching. Furthermore, it has been reported by de Vries and Kelling (1988) that timing performance is improved by using a dynode signal rather than a signal from the anode. There is considerable scope in the choice of the load resistor, and the same considerations that apply in choosing the anode resistor are relevant here: typically, $10 \text{ k}\Omega < R_L < 100 \text{ k}\Omega$. If signal transmission is via a matched coaxial cable (invariably, 50Ω) between the PMT and an oscilloscope or some other electronics, matching at the PMT is achieved in Fig. 14.8(b) by choosing $R_S = z_0$, where z_0 is the characteristic impedance of the cable. Matching may be unnecessary or undesirable where signals are of limited bandwidth, say, less than 10 MHz. It is feasible to derive extra voltage gain simply by increasing R_L , but at the expense of reduced bandwidth. Characteristic impedance matching is inappropriate with charge-sensitive preamplifiers; in this case, it is advisable to choose R_L in excess of $100 \text{ k}\Omega$, as small-value resistors impair performance. Whatever the value of R_L , it is always recommended that some finite value for R_S should be adopted, say, $10 \text{ M}\Omega$, as a safety precaution. It is important to decouple the non-signal end of R_L directly to the ground via C_3 , as illustrated, even if the standard series configuration of capacitors is already in place; failure to do this leads to signal distortion, similar in origin to that shown in Fig. 14.12(a) and (b). A voltage divider proposed in Chapter 13 (shown in Fig. 13.30(b)) is based on two power supplies, allowing direct coupling of the dynode signal.

14.8 Passive circuit analysis

The initial analysis concerns operation of a PMT connected to passive components—mainly, resistors and capacitors. The primary signal is one that is initiated by a single photoelectron, amplified by the multiplier, and delivered to the anode. The equivalent circuit of a current generator makes the PMT particularly versatile, depending on the way in which it is configured by external circuitry. There are broadly two options. The first option is to utilize the high-output impedance property by connecting the anode to a load resistor of high ohmic value. In this configuration, the signature of the voltage output does not follow the profile of the anode current. The second option is to use a low-impedance arrangement that is suitable for the transmission of a pulse down a matched coaxial cable. In this case, the temporal profile and the magnitude of the light signal are retained. The relevance of the forthcoming circuit analysis is that the formulations

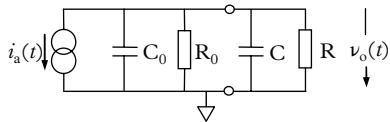


Fig. 14.9. Complete circuit for DC applications of PMTs, based on the equivalent circuit shown in Fig. 14.5(a).

developed realistically describe actual waveforms and signal amplitudes that would be seen on an oscilloscope display.

14.8.1 Anode DC-coupled $R \parallel C$ network

The equivalent circuit shown in Fig. 14.9, configured for negative HV, is adequate for investigating most PMT applications; R and C may be chosen to interface with:

- an oscilloscope probe with a typical impedance of $1\text{ M}\Omega \parallel 10\text{ pF}$
- a length of coaxial cable terminating in a resistor with a resistance R , where R is chosen to be much greater than z_0 , the characteristic impedance (an unterminated cable presents a capacitance of up to 100 pF/m , depending on the cable type selected)
- the input to additional electronics circuitry, such as preamplifiers and amplifiers

We seek an expression for the output voltage, $v_o(t)$, at the anode, generated by the input excitation, $i_a(t)$, which was defined in (14.1). There are several methods for analysing this type of circuit, but the author favours the elegance of the Laplace transform. Without loss of generality, we combine the resistances of the two resistors as $R = R_0R/(R_0 + R)$, and the values of the capacitors as $C + C_0$, in which case the current $i_a(t)$ in Fig. 14.9 can be divided as follows:

$$i_a(t) = i_{C+C_0}(t) + i_R(t) \quad (14.3)$$

so that

$$\frac{-Q_a}{\tau_i} \exp(-t/\tau_i) = (C_0 + C) \frac{d\nu_o}{dt}(t) + \frac{1}{R} \nu_o(t).$$

Taking the Laplace transforms, with $\tau = R(C_0 + C)$, leads to

$$\begin{aligned} \frac{-Q_a}{\tau_i} \frac{R}{(s + 1/\tau_i)} &= L\left(\frac{d\nu_o(t)}{dt}\right)\tau + L\{\nu_o(t)\} \\ &= \tau s L\{\nu_o(t)\} + L\{\nu_o(t)\}, \end{aligned}$$

where $s = j\omega$; therefore

$$L\{\nu_o(t)\} = \frac{-Q_a R}{\tau \tau_i (s + 1/\tau_i)(s + 1/\tau)}.$$

This is a standard form, with an inverse of L^{-1} :

$$\nu_o(t) = \frac{-Q_a \tau}{(C + C_0)(\tau - \tau_i)} \left(\exp(-t/\tau) - \exp(-t/\tau_i) \right). \quad (14.4)$$

For the special case where $\tau = \tau_i$, it is necessary to take limits, using l'Hôpital's rule in (14.4) and arriving at

$$\nu_o(t) = \frac{-Q_a}{(C + C_0)} (t/\tau) \exp(-t/\tau) = \frac{-Q_a}{(C + C_0)} (t/\tau_i) \exp(-t/\tau_i). \quad (14.5)$$

There are three special cases to consider. First, when $\tau \ll \tau_i$, $\exp(-t/\tau) \rightarrow 0$, and $\tau - \tau_i$ is effectively the same as $-\tau_i$. The output signal follows the transient form of the input, (14.1). This is clear in Fig. 14.10, by comparison of the input stimulus (a) with the curve marked $C = 0 \text{ pF}$. Second, when $\tau = \tau_i$ as in (14.5). Third, when $\tau \gg \tau_i$ as in (14.8).

$$\nu_o(t) = \frac{-Q_a R}{\tau_i} \exp(-t/\tau_i). \quad (14.6)$$

The peak signal is

$$\nu_{\max} = \frac{-Q_a R}{\tau_i}. \quad (14.7)$$

The complimentary relationship, when $\tau > \tau_i$ is

$$\nu_o(t) = \frac{-Q_a}{(C + C_0)} \left(1 - \exp(-t/\tau_i) \right), \quad (14.8)$$

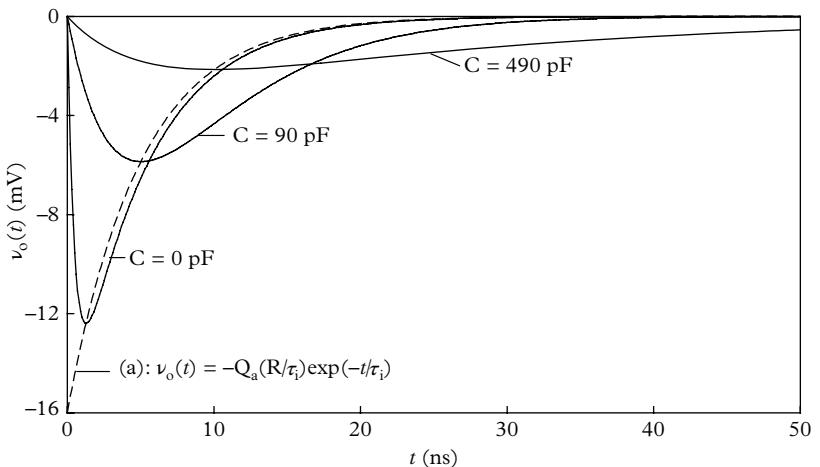


Fig. 14.10. Equation (14.4) with $C_0 = 10 \text{ pF}$; $R = 50 \Omega$; and $Q_a = e\phi = 1.6 \text{ pC}$; this figure illustrates the role of added capacitance, C . The shape of the signal is retained when $\tau \ll \tau_i$; at the other extreme, when $\tau \gg \tau_i$, the output signal exhibits integration, with a slow return to the baseline.

with

$$\nu_{max} = \frac{-Q_a}{(C + C_0)}, \quad (14.9)$$

indicating full integration of the input signal.

14.8.2 Anode AC-coupled R||C network

The previous section concerned analysis of a parallel RC combination, connected between the anode and ground. Assuming the same form for the PMT output, we examine the effect of AC coupling the load resistor to the anode in the manner of Fig. 14.5(b). The circuit configuration, in which an external resistor, R , is AC coupled to the PMT, is given in the equivalent circuit of Fig. 14.11. In practice R_L ranges from 50Ω to $1M\Omega$ and it is always considerably smaller than that resistance of R_0 . In any case, we may absorb R_0 into R_L as a parallel combination as before, but R remains isolated because of C_c . Adopting the previous procedure that led to (14.4), we arrive at the following expression for $\nu_o(t)$, which is manifestly more complex:

$$\nu_o(t) = Q_a \lambda_i / C_0 [K_1 \lambda_i \exp(-\lambda_i t) + K_2 a \exp(-at) + K_3 b \exp(-bt)], \quad (14.10)$$

where the constants are:

$$\lambda_i = 1/\tau_i$$

$$y = \frac{(R_L C_0 + R_L C_c + R C_c)}{R_L C_0 R C_c}$$

$$z = 1/(R_L C_0 R C_c)$$

$$a = \frac{y + (y^2 - 4z)^{1/2}}{2}$$

$$b = \frac{y - (y^2 - 4z)^{1/2}}{2}$$

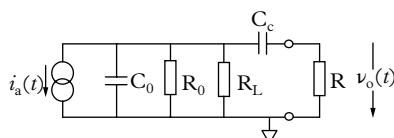


Fig. 14.11. Complete circuit for AC-coupled applications of PMTs, based on the equivalent circuit shown in Fig. 14.5(b).

$$K_1 = \frac{1}{(\lambda_i - a)(\lambda_i - b)}$$

$$K_2 = \frac{-1}{(\lambda_i - a)(a - b)}$$

$$K_3 = \frac{1}{(\lambda_i - b)(a - b)}$$

We note in (14.10) that three time constants are involved: one for the input excitation, and the other two for the capacitors in the circuit. The time constants have been chosen to illustrate the well-known and undesirable effect of baseline shift. The input stimulus is unipolar, but the capacitor, C_c , cannot pass direct current and hence the area of the signal above the baseline is balanced by an equal area below. This law applies whatever the magnitude of the capacitance but by choosing C_c to be large, say, 10 nF, the overshoot becomes hardly discernible, although prolonged. Small capacitor values result in considerable but short-lived overshoot, illustrated in Fig. 14.12 in Curve (b). The user therefore may choose either a small-magnitude but long-lived overshoot, or a peaked overshoot of short duration. Baseline shift produces a particularly unwelcome distortion on both timing and energy resolution, especially where events are randomly distributed in time. The shift varies in proportion to the rate of events

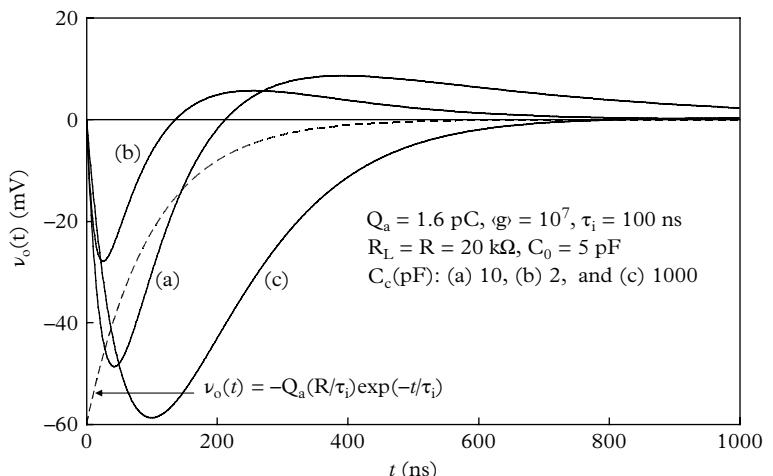


Fig. 14.12. Overshoot effects caused by AC coupling. Curves (a) and (b) show considerable, but comparatively short lived, overshoot. Although present in Curve (c), the overshoot effect is indiscernible, and the shape of this signal closely follows the prediction of (14.6).

and on the mean amplitude of the signals, with the following consequences: the effective threshold set by a pulse height discriminator, for example, will increase with count rate as the height of the pulses, with respect to zero level, decreases. Similarly, for energy encoding, the pulse height or pulse area is progressively undersized with increasing rate. As a check on the validity of (14.10), when C_c is sufficiently large, as is usually the case, z and hence b tend to zero, and the third term in (14.10) vanishes, leading to the expression in (14.4). In practice, C_c is typically 1 to 10 nF, and the resulting pulse shapes are essentially the same as those observed for DC coupling. The overshoot is so small that it is barely discernible in the tail of the pulse shown by Curve (c) in Fig. 14.12, whereas it becomes apparent once the rate is increased. There is a means for preventing overshoot by using passive circuitry, but this will be covered in 14.9.3.

The way to convert a charge pulse to a voltage signal has been presented. Stray capacitance is always present and hence the output waveform is actually developed across an $R \parallel C$ combination. This lengthens and hence distorts the PMT output, as shown in Fig. 14.12, but the effect can be minimized by keeping the values of C and R small, in which case, the output-voltage signal will closely follow the time signature of the input light excitation. There are applications in which this is important, for example, fluorescence decay, photon counting at high signal rates, and fast timing. Conversely, there are other applications in which preserving the original excitation time signature is unnecessary, and possibly undesirable. Where the aim is to measure the charge content of a pulse, to relate it to the energy of a radionuclide, for example, smoothing and stretching has the effect of reducing noise while benefiting charge determination; if sufficient integration is applied, the peak of each pulse becomes proportional to its area, and thus ideal conditions for analogue-to-digital conversion are obtained. What should now be apparent is that signal handling for best timing fidelity, and that for optimal energy encoding, are quite different and, if both are required, one of the schemes shown in Fig. 14.8 must be adopted.

14.9 Signal-shaping preamplifiers

The idea that a PMT signal can be modified to suit the application has been touched on in the previous section. Predicted waveforms for specific high-pass and low-pass RC combinations are readily confirmed by capturing signals on an oscilloscope. However, agreement between analysis and theory only holds provided CR and RC circuits are unloaded, as measured, for example, with a high-impedance oscilloscope probe. The requirement to shape signals without loading the electronics, to provide gain and a low-output impedance drive, is achieved using buffer amplifiers with high-input and low-output impedance.

14.9.1 Charge-sensitive preamplifiers

A charge-sensitive preamplifier is a variant of the generic preamplifier shown in Fig. 14.1, in which $R_f \rightarrow \infty$. It produces a voltage step, followed by a slow decay, at its output. This signal is ‘differentiated’ by the next stage to reduce the decay time to microseconds; there is generally no loss in resolution by narrowing the pulse in this way because the charge information resides primarily in the rise time of the output pulse in Fig. 14.13. Charge-sensitive preamplifiers are therefore the preferred choice for the input module in spectroscopy applications. The analysis of the circuit depicted in Fig. 14.13 follows readily by ignoring R_f . This is permissible because R_f is, in practice, very large in magnitude.

The following relationships follow from the figure:

$$v_o = -Av_i$$

$$v_f = (1 + A)v_i$$

$$Q_f = C_f v_f = C_f (1 + A) v_i$$

If we assume that the amplifier has high open loop gain and high-input impedance, then $Q_a = Q_f$, and the dynamic input capacitance of the amplifier, C_{in} , is

$$C_{in} = Q_a/v_i = Q_f/v_i = C_f(1 + A).$$

The effective, or total, capacitance at the input is extremely high, with

$$C_{eff} = C_0 + C_{in} = C_0 + C_f(1 + A) \approx AC_f, \quad (14.11)$$

where the assumption is that C_0 is negligible compared with the product AC_f . A preamplifier such as the ORTEC 142A has $A = 40,000$, with C_{eff} in the nanofarad range. The conversion gain in volts per coulomb is

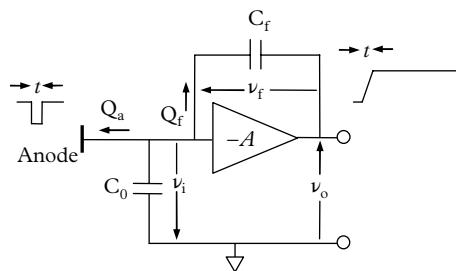


Fig. 14.13. Schematic for the input section of a charge-sensitive preamplifier of open loop gain $-A$; C_0 represents the capacitance of the detector plus that of the leads. In practice, a resistor, R_f , of resistance up to $1\text{ G}\Omega$, is connected across C_f to discharge it between events. Note how the input signal is integrated.

$$A_Q = v_o/Q_a = -Av_i/(C_{\text{eff}}v_i) = -1/C_f \quad (14.12)$$

and

$$v_o = -Q_a/C_f. \quad (14.13)$$

It is important to understand the dynamic behaviour of the circuit shown in Fig. 14.13 when it is presented with a pulse from a PMT. As the capacitor C_f progressively integrates this charge, v_o increases at a rate comparable to the fwhm of the input signal. In the ideal circuit shown in Fig. 14.13, v_o reaches a maximum, where it remains once all the charge has been collected. The analysis so far presents a way of shaping anode signals with a preamplifier, designed for the precise measurement of charge. A charge-sensitive preamplifier also has the further important attribute of low-output impedance, typically 50 or 93 Ω . This means that the preamplifier can drive a matched coaxial cable or successfully interface to a shaping amplifier with typically 1 k Ω input impedance. It is important to appreciate that a charge-sensitive amplifier does not amplify charge; instead, it absorbs the total charge in an input pulse on a small capacitor, C_f . This creates a significant voltage signal, the peak amplitude of which is proportional to the total input charge.

It should be emphasized that (14.11) to (14.13) only apply subject to $A \gg 1$ over a specified frequency range; C_0 must be small compared with AC_f . Provided A is sufficiently large in (14.11), then $C_f(1 + A) \gg C_0$ can be satisfied even if $C_f \ll C_0$. This is important because, subject to satisfying these conditions, the output voltage from (14.13) will vary as $-Q_a/C_f$. With PMTs, C_0 is typically 20 pF, and preamplifiers with $C_f = 1$ pF are commonly used. In older texts, C_0 is referred to as cold capacitance; as the term suggests, this is the capacitance at the preamplifier input when the circuit is inactive: it comprises the output capacitance of the PMT and its voltage divider, that of the coaxial lead used to connect the preamplifier to the PMT, and any stray capacitance; C_{eff} is the dynamic capacitance and it is dominated by the feedback component. The sensitivity, A_Q , is essentially independent of C_0 , from (14.12), and hence stable with changes in C_0 . This means that the conversion gain (charge to voltage) is largely independent of the length of the input coaxial cable—a highly desirable attribute for remote locations, or where microphony is a consideration.

Although amplifier operation is relatively insensitive to input capacitance, for optimal performance it is nonetheless beneficial to minimize the total cold capacitance at the preamplifier input. There is an approximate 20 pF capacitance associated with the anode of a PMT once it is inserted into a socket; about half the capacitance is contributed by the socket and its dielectric, with the remainder contributed by the close proximity of the anode mesh to the last dynode. Users have no control over this capacitance but the contribution from the socket may be reduced by isolating the anode, effectively by keeping it remote from the socket using a flying lead connected directly to the preamplifier. This is only recommended for attaining the very lowest noise with the highest-quality preamplifier. The choice of connecting

leads is important. Manufacturers prefer RG62U ($93\ \Omega$), because its capacitance is only $44\ pF/m$, compared to that of $100\ pF/m$ for the ubiquitous $50\ \Omega$ RG174/U. RG62U has the further advantage over $50\ \Omega$ cable of almost doubling ν_o . The author is unclear as to the reason for the restricted use of RG62U when it has the advantage of delivering twice the signal of RG174/U. Performance is enhanced by mounting the preamplifier on or close to the PMT socket, thereby avoiding the use of an input coaxial lead; the compact Amptek Inc. A200 series is ideal in this respect. The noise N , incurred by adding capacitance, is specified by dN/dC and is typically 5 to 10 electrons/pF. Incremental capacitance has the most marked effect on those preamplifiers with an intrinsically low N value. The subject of noise is covered in 14.10.2.

14.9.2 Commercial charge-sensitive amplifiers

The rise time of charge-sensitive preamplifiers is usually fast: for the ORTEC model 142A, for example, it is 5 ns for no added capacitance, and 12 ns for $C_0 = 100\ pF$. Consequently, for fast pulses, the integrating action of a charge-sensitive preamplifier leads to an output with a relatively fast rise time and a long decay set by the time constant $R_f C_f$ (see Fig. 14.14). For slow pulses, such as those from a NaI(Tl) scintillators, the preamplifier rise time will follow the decay time of the pulse in the manner of (14.4) or (14.10).

Another desirable property of charge-sensitive amplifiers is that the integrating action be independent of the time structure of the input pulses. Consequently, pulses with different rise and fall times, but with the same charge content, produce

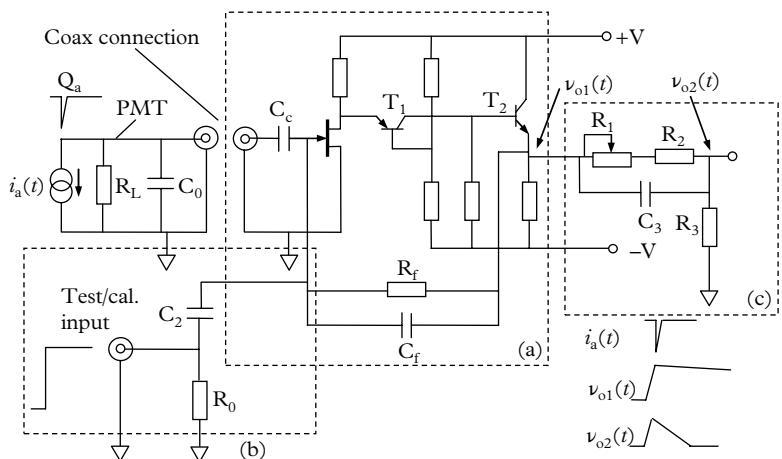


Fig. 14.14. Circuit diagram of a typical charge-sensitive preamplifier adapted from Herbst (1970). The rise time of the output pulse is approximately equal to the width of the input pulse Q_a . Connection of the PMT to the preamplifier input, C_c , is a length of unterminated coaxial cable of characteristic impedance, z_0 (not shown).

the same peak height—this is precisely the requirement of an energy-encoding circuit. The integrating action is particularly important where photoelectron numbers are low. Historically, modular detector systems suitable for PMTs are those that were designed for nuclear spectroscopy (e.g. NaI(Tl) and plastic scintillators) but the argument put forward in this section shows that they are also appropriate for fast single- or multi-photon signals.

Until the late 1960s, the preferred input device for a charge-sensitive amplifier was a radio valve. The use of this relic may appear puzzling, since the silicon junction transistors available at this time already offered high-frequency performance. The explanation for the valve's long survival lies in its noise performance, which was ultimately surpassed by the availability of FETs. It is interesting that, 50 years later, the same FET-based circuit design appears in today's high-sensitivity and low-noise-performance preamplifiers. Furthermore, despite the advances in FET technology, JFETs are still preferred over MOSFETs because of their superior noise performance. According to Radeka (1988), the series noise of a MOSFET is some 2 to 6 times higher than that of a JFET. Figure 14.14 shows the essential elements of classical charge-sensitive preamplifiers. Full circuits are given by Gillespie (1970) and by Mori (1996), for example.

Figure 14.14 shows an idealized charge-sensitive preamplifier, which is divided into three sections (indicated by the dashed lines) for ease of explanation. In Fig. 14.14(a), which shows the charge-sensitive transimpedance section, the FET and T_1 form a high-gain cascode amplifier feeding an emitter follower, T_2 , which provides low-output impedance to drive the next section of the preamplifier. The feedback path $R_f \parallel C_f$ between output and input performs charge-to-voltage conversion. R_f ranges from megohms to gigohms in commercial circuits and, consequently, the conversion from charge to voltage is determined solely by C_f . It follows from the basic relationship $Q = C_f V_o$ that a 1 pF capacitor will provide a signal output of 1 V/pC of charge input. To put this into context, according to (14.13), a single-electron event from a PMT operated at 10^6 gain will produce a voltage pulse of 160 mV amplitude across a 1 pF C_f . It is a worthwhile exercise to substitute different parameter values to examine the effect on the output signal.

The resistor R_f serves two purposes. First, it provides low-frequency feedback stability; second, it partially discharges C_f after each pulse. Typical values for the resistance R_f and the capacitance C_f are $100 \text{ M}\Omega < R_f < 1 \text{ G}\Omega$, and $1 \text{ pF} < C_f < 20 \text{ pF}$, respectively, giving a correspondingly long discharge time constant T of $100 \mu\text{s} < T < 1 \text{ ms}$. In the charge-sensitive mode of operation, Johnson noise manifests itself as a current source $dI^2 = 4kTdf/R_f$ at the input, and hence R_f should be large to minimize this contribution. For a detailed discussion of this noise source, see Section 5.3.2 in the work by Nicholson (1974). Omission of R_f , or choosing R_f too big, causes the voltage train at the preamplifier output to progress towards the positive rail, even for modest signal rates, ultimately cutting off the preamplifier output. Even with R_f in place, it is obvious that some pulses overlap one another, but the effect can be limited by the addition of the shaping section, shown in

Fig. 14.14(c), which reduces the pulse decay time by a factor of 10. The action of R_f is referred to as ‘dynamic charge restoration’ but, in an ingenious circuit due to Madden *et al.* (1969), R_f is omitted and replaced by a pulsed optical-feedback arrangement to discharge a feedback capacitor of only 0.2 pF; R_f is also omitted in the ‘pulsed transistor reset’ circuit of Radeka (1970) and Landis *et al.* (1998). These ultra-low-noise preamplifiers, with noise figures approaching 30 ENC_s, are ideal for Si and Ge detectors operating at liquid nitrogen temperatures, but their performance is far in excess of that required by PMT applications. The A203 standalone cooled preamplifier from Amptek, for example, has a noise figure, ENC, of 20 electrons. The A225, from the same manufacturer, (uncooled), has noise of 280 ENC_s. This topic is discussed in detail in later sections.

14.9.3 Pole-zero cancellation

Referring to Fig. 14.14, coupling a charge-sensitive preamplifier and a PMT requires a capacitor C_c to prevent DC voltage amplification, which otherwise results in a DC gain of R_f/R_L , where R_L is the load resistor. R_f is chosen to be orders of magnitude higher than R_L , to derive the benefit of charge-sensitive detection, and any input-voltage offset is magnified by this factor. Including a capacitor, C_c , removes this particular problem, but introduces another. As is evident in Fig. 14.12, AC coupling causes overshoot and consequently baseline shift with count rate. However, undershoot that would otherwise appear at the output can be removed by an ingenious passive circuit devised by Nowlin and Blankenship (1965) and shown in Fig. 14.14(c). Their technique consists of shunting C_3 with a resistor $R_x = R_1 + R_2$ and then choosing the time constant $C_3R_x = C_f R_f$. The output decays as $C_3[R_3R_x/(R_3 + R_x)]$ and, if $R_x \gg R_3$, v_{01} will follow a time constant close to C_3R_3 , but without undershoot. In addition, with this circuit we are able to decrease the inordinately long time constant $C_f R_f$ by a factor of 10 or more, without having to reduce R_f . It is helpful to pursue these considerations, referring to a typical low-cost preamplifier described by the following parameters:

$$R_f = 2.2 \text{ M}\Omega$$

$$C_f = 220 \text{ pF}$$

$$R_x = 22 \text{ k}\Omega$$

$$C_3 = 22 \text{ nF}$$

$$R_3 = 2.4 \text{ k}\Omega$$

$$\text{the primary time constant } C_f R_f = C_3 R_x = 484 \mu\text{s}$$

$$\text{the secondary time constant } C_3[R_3 R_x / (R_3 + R_x)] = 47.6 \mu\text{s}$$

It should be clear that there are many sets of component values that will provide satisfactory performance in Fig. 14.14.

Pole-zero (p/z) cancellation reduces the long tail seen in ν_{02} by a factor of 10, in addition to removing the undershoot. These shaped pulses are still needlessly long since, through the integration performed by C_f , the charge content is captured in the first ~ 10 ns for a fast plastic scintillator and, correspondingly, 1 μs for much slower NaI(Tl), for example. It is therefore customary to further reduce the output pulses to a width of about one μs . This is one of the functions of a shaping amplifier. An output buffer (not shown in Fig. 14.14) is necessary after a p/z circuit, usually consisting of a differential pair feeding a push-pull arrangement of emitter followers; this provides low-output impedance, typically matched to the coaxial line. Note that p/z baseline restoration is included in shaping amplifiers and other types of amplifiers.

14.9.4 Preamplifier calibration

One major attribute of charge-sensitive preamplifiers is the ease with which they can be calibrated for charge sensitivity; this is accomplished by injecting a voltage step from a precision pulser into the test input shown in Fig. 14.14(b). If the capacitance is, say, 33 pF, a 1 V step would then inject 33 pC into the preamplifier, producing a measurable voltage signal at the output. This may be viewed, for example, on an oscilloscope or connected via a shaping amplifier to an MCA. The key point to appreciate is that, whatever instrumentation may follow the preamplifier, we can relate the final output, in volts or channel number, to a known input charge. Knowledge of any internal amplification that may apply within the preamplifier or shaping amplifier is not required. The calibration facility provides another way in which to determine PMT gain based on a single-photoelectron source. Applying a step voltage leads to a calibration of the system in terms of picocoulombs per channel and, if the peak of the SER appears in a particular channel, we immediately know the most probable multiplier gain, and the mean gain $\langle g \rangle$ is readily calculated from the MCA spectrum of pulse heights. This is fully discussed in 5.8 and 5.9. Furthermore, a calibrated system leads to the absolute determination of scintillator efficiency in terms of photoelectrons per keV, or photons per keV for a calibrated photocathode.

14.10 Signal-shaping main amplifiers

14.10.1 Signal-shaping theory

Figure 14.15 shows an example of a modular detector system for scintillator spectroscopy. Shaping, described in 14.9.2, is first applied in the preamplifier to generate a pulse with a decay time constant of the order of 50 μs . But, further shaping is performed in a main amplifier to optimize three important parameters:

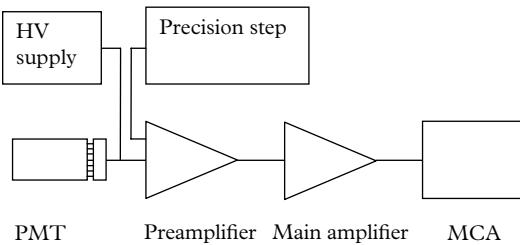


Fig. 14.15. The elements of a spectroscopy system assembled from commercial units.

resolution, noise, and count-rate capability. The analysis now adopted is restricted to a basic level, but is sufficient as an introduction to pulse shaping.

Circuits used to accomplish pulse shaping are based on combinations of CR and RC networks, referred to as differentiating and integrating circuits, respectively; inductors are, however, seldom used. Readers may be puzzled by the rationale for a combination of two electronic units that would appear to leave the signal unchanged. The stated mathematical description, in the case of a CR network, for example, is realized only for CR sufficiently small and, conversely, a network will integrate provided that RC is large compared with the duration of the input signal, but note the mathematical terminology tends to be used whatever time constants apply.

Commercial main amplifiers are available offering various combinations of CR and RC circuits. The general designation used for these amplifiers is $(CR)^n(RC)^m$, where n and m are integers indicating the total number of shapers, $n + m$. The transient behaviour of the simplest filter combination, $(CR)(RC)$, is shown in Fig. 14.16. This amplifier is analysed first, followed by a filter configuration with two integrators, $(CR)(RC)^2$ (see Fig. 14.17). An ideal charge-sensitive amplifier, following the analysis of 14.9.1, is assumed at the input; it integrates each PMT pulse producing a voltage step of Q_a/C_0 . The output of the charge-sensitive preamplifier is the source for the C_1R_1 differentiator, with $v_1(t) = Q_a/C_0 \exp(-t/\tau_1)$. The output signal of the RC integrator contains two exponential terms with time constants of τ_1 and τ_2 , respectively, derived using an analysis similar to that in 14.8.1. The Laplace transform for the three elements in Fig. 14.16 is

$$\frac{1}{s} \times \frac{s\tau_1}{(1+s\tau_1)(1+s\tau_2)} = \frac{\tau_1}{(1+s\tau_1)(1+s\tau_2)},$$

with the inverse

$$v_2(t) = \frac{Q_a}{C_0(\tau_1 - \tau_2)} \left(\exp(-t/\tau_1) - \exp(-t/\tau_2) \right).$$

Sometimes it is expedient to replace $1/\tau_1$ and $1/\tau_2$ by λ_1 and λ_2 , respectively, to yield more compact expressions such as the following:

$$\nu_2(t) = \frac{Q_a}{C_o} \frac{\lambda_2}{(\lambda_2 - \lambda_1)} \left(\exp(-t\lambda_1) - \exp(-t\lambda_2) \right). \quad (14.14)$$

A special case of wide interest is the adoption of common parameter values for $\lambda_1, \dots, \lambda_n$ generating the asymptotic form of $\nu_2(t)$. The required expression is derived using l'Hôpital's procedure with respect to λ_1 , to obtain

$$\nu_2(t) = (\lambda t) \exp(-\lambda t).$$

By adding a second integrator, $\tau_3 = R_3 C_3$, the Laplace transform takes the form $\tau_1/(1+s\tau_1)(1+s\tau_2)(1+s\tau_3)$ leading to (14.15)

$$\nu_3(t) = \frac{Q_a \lambda_2 \lambda_3}{C_0} \frac{(\lambda_3 - \lambda_2) \exp(-\lambda_1 t) + (\lambda_1 - \lambda_3) \exp(-\lambda_2 t) + (\lambda_2 - \lambda_1) \exp(-\lambda_3 t)}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_2)(\lambda_1 - \lambda_3)}. \quad (14.15)$$

The special case of $\tau_1 = \tau_2 = \tau_3 = \tau$ is realized by using l'Hôpital's rule with respect to λ_1 , to obtain

$$\nu_3(t) = \frac{(t/\tau)^2}{2!} \exp(-t/\tau)$$

and, generally,

$$\nu_n(t) = \frac{(t/\tau)^{n-1}}{(n-1)!} \exp(-t/\tau). \quad (14.16)$$

Adding integrating stages is a folding process and hence (14.16) takes the form of the familiar factorial function, also encountered in scaling (see e.g. 4.4 and 4.18.2). Note, for n large, this expression tends towards a normal distribution—hence the expression Gaussian filter.

Commercial main amplifiers usually include a bipolar signal output, achieved by differentiating (14.16) and shown in Fig. 14.18 for $\nu_2(t)$. As for the parent, the bipolar signal becomes more symmetrical as the number of filter stages is

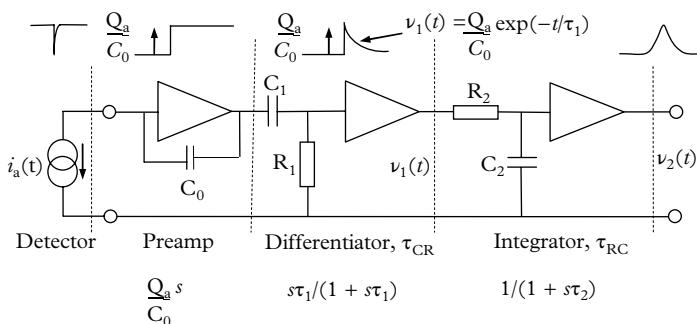


Fig. 14.16. Schematic of a shaping amplifier of type (CR)(RC). Ideal unity-gain-buffer amplifiers isolate individual stages from one another. Time constants are given by $\tau_1 = C_1 R_1$, and $\tau_2 = R_2 C_2$. The complex variable s refers to $j\omega$.

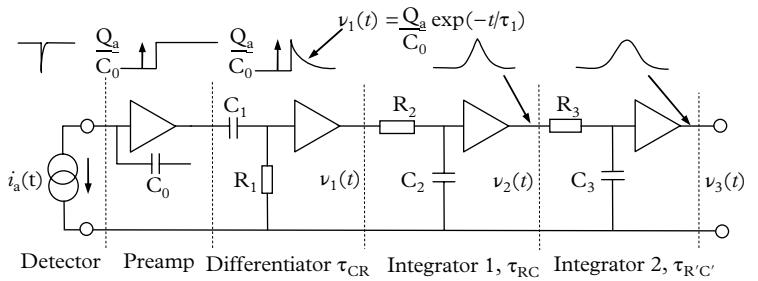


Fig. 14.17. Schematic of a shaping amplifier of type (CR)(RC)². Ideal buffer amplifiers, of unity gain, isolate individual stages.

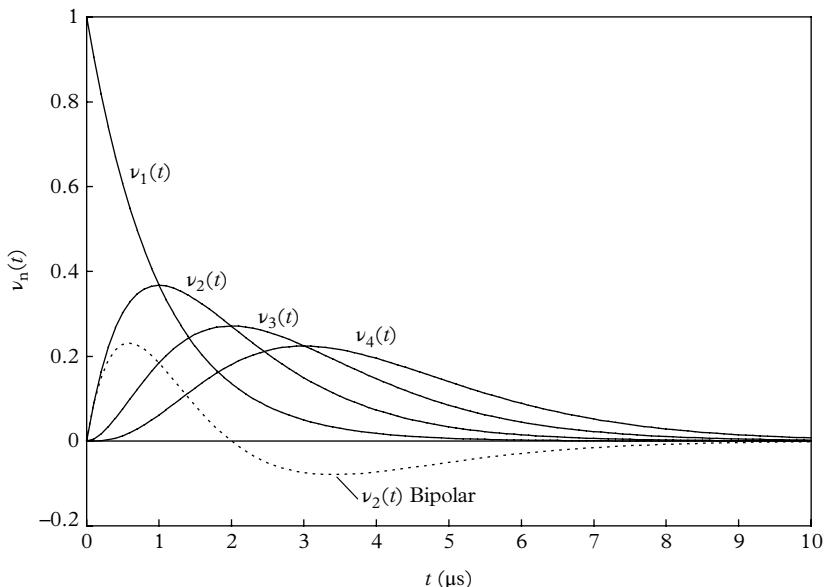


Fig. 14.18. A special case by adopting a common time constant of 1 μs for all modules illustrates how the input is progressively smoothed, flattening the peak of the input exponential, as required. The bipolar waveform is derived from (14.16).

increased. The waveforms in Fig. 14.18 reveal an effect known as the ballistic deficit. The signal $v_1(t)$ represents a pulse of decay time 1 μs, and $v_2(t)$ is a shaped signal derived from adding a single integrator circuit. Taking the peak height of $v_2(t)$ as a measure of the size of the input pulse, it is clear from Fig. 14.18 that this is derived from the collection of only two-thirds of the PMT output. However, the output after three stages of integration, $v_4(t)$, corresponds to 95% collection and, consequently, to optimal resolution on statistical grounds. The deficit can be reduced by increasing the RC time constants, or by increasing the number of

stages, but at the expense of the count-rate capability—the usual trade-off. Generally, the adoption of a common time constant gives the best combination of flat top and minimal overlap. Pulse height resolution is degraded by the bipolar process in favour of reduced dead time and baseline stability. The time taken for the output pulse to reach its peak is given by $n\tau$, known as the peaking time, τ_p . Response curves become more symmetrical with increasing n , ultimately becoming Gaussian. There is an obvious trade-off between pulse height resolution and count-rate capability: additional integrators reduce noise, but the output pulse is progressively stretched. Improved performance, in the presence of high and variable count rates may also be achieved by selecting $(CR)^2$ -RC or CR-RC-CR configurations. The additional differentiation, if suitably tuned, can provide a reasonably symmetrical bipolar pulse. The circuits described are for illustrative purposes without mention of gain provision. A typical commercial NIM unit offers continuously selectable voltage gain of up to $\times 3000$, with incremental shaping adjustment from 0.1 to 10 μs , and the choice of unipolar and bipolar pulse outputs.

14.10.2 Detector noise

Recall that the PMT traces in Figs 8.17 and 8.18, which illustrate scintillator detection, show considerable fine structure, even for signals corresponding to 1000 pe. This noise is a limiting factor in spectroscopy because of its effect on resolution. Integrating an anode pulse on a capacitor is the best means for determining its size. This is essentially what a charge-sensitive preamplifier does but a practical detector system must also include a shaping amplifier. With NaI(Tl), for example, collection is complete after 1 μs (four time constants), and hence a circuit that encodes the pulse over this timescale is appropriate. The capacitor in the charge-sensitive preamplifier must be discharged before the next event, by including a parallel resistor R_f . Capacitors do not produce noise—only resistors do—and this is one of the features of charge-sensitive preamplifiers. Those with the lowest noise include a JFET input. Detector noise is covered in the review paper by Radeka (1988); this review is highly recommended for those seeking an erudite treatment of the subject. The consideration here follows the analysis given by Gillespie (1970) for a detector consisting of a charge-sensitive preamplifier followed by differentiating and integrating shaping amplifiers (see Fig. 14.19).

We consider the standard equivalent circuit for a PMT, in which C_0 is the total cold capacitance due to connecting cables, the PMT anode configuration, and stray capacitance. The PMT is connected to a charge-sensitive preamplifier, the output of which feeds a differentiator and an integrator in the manner shown in Fig. 14.16. The noise, in units of rms volts, from the major sources v_s , v_f , and v_g combine in quadrature to give a total noise variance of $v_s^2 + v_f^2 + v_g^2$. Gillespie gives expressions for the three contributions in terms of two independent time constants, τ_{CR} and τ_{RC} , and states that the noise is minimal when the two are equal, which is assumed henceforth. The symbols and their definitions are given as follows:

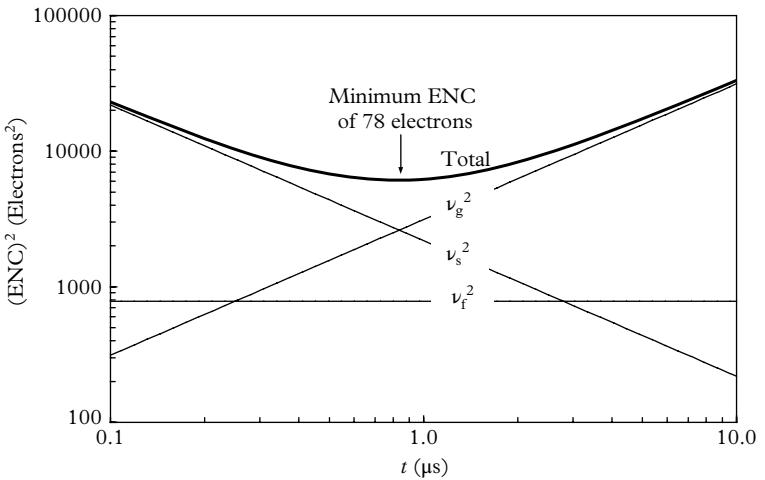


Fig. 14.19. Individual contributions to the total noise from sources discussed in the text. The position of the minimum is referred to as the peaking time, τ_p .

k is the Boltzmann constant, 1.38×10^{-23} joule/K

T is the absolute temperature, taken as 290 K

g_m is the transconductance of the JFET, 10 mA/V

τ is the shaping time and, for illustrative purposes, $\tau_{CR} = \tau_{RC} = \tau$

C_0 is the total cold capacitance, defined previously

I_g is the gate current

$\nu_s^2 = 0.7/g_m \times kT/2\tau$ is the shot noise associated with the current flowing in an FET channel; this is expressed as an equivalent noise resistance, R_n , which if connected to the gate of the JFET would generate thermal noise equal to that in the FET channel; for a JFET, this equivalent resistance is given by $0.7/g_m = R_n$

$\nu_f^2 = 10^{-13}/2$ is the contribution from $1/f$ noise

$\nu_g^2 = eI_g\tau/4C_0^2$ is the JFET gate current shot noise, $\sim 2 \times 10^{-9}$ A

Noise is expressed in terms of the equivalent charge, ENC, in coulombs or, more appropriately, in electrons. This follows by dividing the noise charge by the electronic charge, as in (14.17), to provide a standard specification for use in charge-sensitive applications. The variance in terms of $(ENC)^2$ is given by

$$(ENC)^2 = (C_0/e)^2(\nu_s^2 + \nu_f^2 + \nu_g^2). \quad (14.17)$$

The rms noise, in units of electrons, may be deduced from the individual and total curves in Fig. 14.19 by taking the square root of the ordinate; the noise minimum occurs for $\tau \sim 1 \mu\text{s}$, for which the noise is 78 electrons, in this particular

example. Increasing the cold capacitance of the preamplifier, by adding a signal lead and the detector, results in an increment in noise of between 1 and 10 electrons/pF. Note in (14.17) that the contribution from ν_g is independent of C_0 ; ν_s^2 varies as $1/\tau$, and ν_g^2 scales as τ . Charge-sensitive preamplifiers are suitable for use with semiconductor radiation detectors, PMTs, APDs, and ionization chambers. There is provision in some for an HV bias supply for semiconductor detectors but this feature may be ignored if not required.

14.10.3 Modular detection equipment

Shaping amplifiers tend to have a single control switch for selecting the shaping time. There are generally up to six positions (including ‘out’), making for timely selection of the optimal setting. Note, in arriving at this setting, one should examine not only the pedestal noise but also the energy resolution. The more advanced ‘Research Amplifier’ allows independent adjustment of shaping time constants, for both integrating and differentiating: there is a total of 100 shaping combinations (τ_{CR} , τ_{RC}), with both of the ranges extending to 10 μs . Selecting the best pair is not such a daunting task as the number of combinations might suggest: the approximate location of the optimum soon becomes apparent. In any case, the optimum is broad with respect to the choice of settings. Noise generally follows a parabolic-like curve with a minimum corresponding to a shaping time of around 1 μs ; this explains why the range of shaping times is centred on this value. We see the justification in (14.17) for keeping cold capacitance to a minimum. There is a relationship between t_{fwhm} of the anode pulses and τ_p ; about 1 μs is ideal for NaI(Tl), which is intuitively obvious since all but 2% of the charge information from NaI(Tl) is gathered within the first microsecond, and the tail of an event contains noise but little signal contribution. For the much faster pulses from plastic scintillators, τ_p , the optimal setting can be shown to be of the order of 20 ns.

An MCA is ideal for spectroscopy. It allocates a number to the size of every input signal, building up a spectrum of pulse heights as data is accumulated in a memory of typically 1000 contiguous channels. The user can follow data accumulation on a display screen, either free-running or set to a specific run time. The key electronic function of the instrument is one of analogue-to-digital conversion of input signals ranging up to 10 V in a typical instrument. Some versions incorporate a preamplifier and a shaping amplifier, but superior performance is usually obtained with external electronics units selected from the NIM range. The internal signal processing time of the instrument, the dead time, is of the order of 1 to 2 μs . Operating at count rates that incur dead time in excess of 20% should be avoided because the correction is approximate and the recorded spectrum is invariably distorted.

Preamplifiers have preset sensitivity, and the conversion gain of the MCA is also fixed, leaving PMT gain and shaping-amplifier gain as the only variables, in practice. Conversion gain refers to the number of channels that are used to record a spectrum, quoted in powers of 2: 256, 514, 1024, and so on. There is also

provision for selecting the shaping-amplifier time constant, τ_s , as discussed previously. The objective is reduced to one of allocating gain optimally between the PMT and the shaping amplifier, to minimize the effects of noise. Shaping amplifiers have switch-selectable coarse gain ranging from $\times 10$ to $\times 3000$ in six steps, and a choice of τ_s ranging from 0.1 to 10 μs .

There are two aspects to noise in modular spectroscopy. The low-energy region of a recorded spectrum, spanning 1 to 20 channels, invariably contains counts originating from sources other than signal. This group of counts can be excluded from display by setting the lower-level discriminator of the MCA at the noise edge, thereby reducing dead time. However, this changes the dynamic range from 1000:1 to typically 100:1. The objective is to find means for locating the origin of these noise counts, and ways to reduce or eliminate them. This is established by recording a spectrum with the PMT in circuit but with the HV switched off, or by temporarily disabling modules that make up the system. Under these circumstances, any counts that appear in the low-channel region of the spectrum are noise in origin. Noise in low channels is characterized by a steeply falling spectrum, even when displayed on a log scale. In high-energy physics, this noise is referred to as pedestal, but here we will be concerned with defining an effective cut-off at low channels, above which noise is predominantly absent. This is the noise edge and, although subjective in its precise definition, reflecting the uncertainty concerning its exact location, the onset of spurious counts is sufficiently steep that this is not a critical issue. An example illustrating noise edge is given in 12.22. The module primarily responsible for system noise can be exposed by adopting a simple procedure, with the HV off. The noise spectrum displayed by an MCA is observed with the preamplifier output disconnected from the shaping amplifier. A module of any quality should show negligible noise, even in the lowest MCA channels and at the highest gain setting of $\times 3000$. The effects of different shaping time constants are also easily checked at the same time. A second test procedure may be performed with the preamplifier reconnected to the shaping amplifier; the PMT is in circuit, but inactive, thus revealing the noise contribution from the preamplifier and the associated external capacitance alone.

14.11 Fast amplifiers

14.11.1 Discrete-component amplifiers

The availability of fast silicon transistors in the mid-1960s saw a rapid development in amplifiers for pulsed and other applications. A selection of the key circuit configurations is given here, from which most of the commercial pulse amplifiers subsequently evolved. Their performance is outstanding, and a more extensive treatment may be found in Nicholson (1974). A common misconception is that

OP-AMPS, in the form of integrated circuits, would by now have replaced transistorized circuits, in the same way that transistors superseded radio valves. OP-AMPS with a gain-bandwidth product exceeding 1 GHz became available only relatively recently, with the current-feedback types providing a closed loop gain of about 10, extending over a bandwidth of up to 200 MHz. This performance should be compared with that of the HP low-noise silicon transistor AT41533, with an f_t of 10 GHz, and the microwave monolithic integrated circuit (MMIC) Darlington amplifiers offered by Mini-Circuits (2009) and with performance up to 8 GHz. The discussion that follows refers to amplifiers of relatively modest gain that preserve pulse shape over previous considerations, which solely concerned charge measurement. Circuit designs appeared in the scientific literature in the early 1960s following the introduction of silicon transistors that were faster than those available in the late 1950s. The long-tail pair arrangement adopted by Collinge *et al.* (1965) (see Fig. 14.20) is typical of the simplicity of design employed by Cherry and Hooper (1963), Rush (1964), Lunsford (1964), and Millard (1967), as illustrated in Figs 14.20, 14.21, and 14.22. Modules have gain between 5 and 10, with a rise time in the region of 1 ns. Duplicate stages may be cascaded to provide an overall gain of 1000 but with degraded rise time. Performance parameters are quoted as $G = 1580$, and $t_r = 3.05$ ns, by Rush; and $G = 270$, and $t_r = 1.0$ ns, by Millard, for example. The key parameters quoted by Rush were achieved by cascading five stages of the design shown in Fig. 14.21, while Millard used three stages. Even by today's standards, these are impressive performance figures. What is particularly noteworthy in all cases is the adoption of minimal components to achieve the required performance.

An important, although sometimes overlooked characteristic of any amplifier is the way it recovers from overload. In nuclear and high-energy physics detectors, overload may result from the effect of out-of-geometry cosmic ray events, or signals originating from the natural background. The time taken to recover, and the

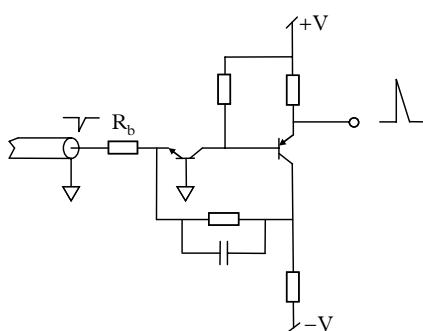


Fig. 14.21. A grounded-base-input amplifier with a gain of 4.5 and rise time of 1.25 ns, due to Rush (1964).

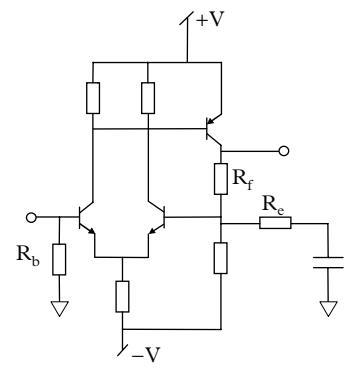


Fig. 14.20. Non-inverting, long-tail pair input with feedback, due to Collinge *et al.* (1965). Gain is given by R_f/R_e .

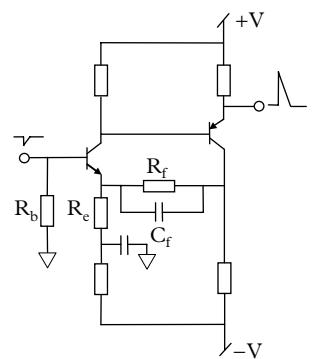


Fig. 14.22. A voltage-series feedback pair of the type developed by Cherry and Hooper. Voltage gain is R_f/R_e but falls off at high frequency as C_f becomes effective. Feedback is of the type shunt output series input.

possible generation of spurious pulses during recovery, are both important. The best amplifiers in this respect are directly coupled ones, which have the further advantage of freedom from baseline shift with rate. The idealized long-tail pair arrangement (also known as an emitter-coupled differential amplifier) shown in Fig. 14.20 has good overload characteristics and a gain of 10. Its symmetrical input has long been appreciated for its high common mode rejection. Input impedance is high for this configuration, allowing the freedom to select the input-biasing resistor, R_b , to match the transmission-cable impedance.

A common base transistor has the desirable property of fast response and low input impedance of $r_e = kT/eI_e \approx 26/I_e$, where I_e is expressed in milliamps. A feedback amplifier in this configuration, due to Rush (1964), is shown in Fig. 14.21. The circuit presents low input impedance and is ideally connected directly to the anode of a PMT (operated at negative HV). When remotely located, the emitter input impedance constitutes only part of the signal cable termination, and a series resistor must be included to achieve matching. More recent versions of this circuit have been presented by Busso *et al.* (1999) and Baturitsky and Dvornikov (1999), with the latter authors offering a grounded-base-and-transimpedance combination in the same circuit. The voltage-series feedback pair shown in Fig. 14.22 has relatively high input impedance; the circuit amplifies the voltage developed across the input bias resistor, R_b . As with the other designs, R_b can be set to achieve matching or, if the amplifier is mounted close to the anode, R_b may be increased to enhance the overall voltage gain. The circuit shown illustrates the principle of operation but, in reality, the biasing arrangement is more complicated than indicated. Closed loop gain is given by R_f/R_e , and cascading is possible, to achieve an overall gain of 100, with a rise time of 2 ns. In addition, C_f is selected within the range 1 to 5 pF to reduce overshoot and to maintain stability; recovery from overload is particularly fast. The circuit of Fig. 14.22 formed the basis of the LeCroy NIM model 333 bipolar amplifier, which was popular in the 1970s. Circuit analysis of this amplifier may be found in the work by Millman and Halkias (1972).

14.11.2 Amplifiers incorporating microwave transistors

A spin-off from mobile phone development is the availability of microwave frequency transistors at very low cost. Amplifiers offering gain of 10 with a cut-off frequency of 5 GHz are the products of most interest to users of the fastest PMTs, particularly microchannel devices. The building block of the ERA series of Darlington amplifiers, shown in Fig. 14.23 and offered by Mini-Circuits and Avantek, is the MMIC fast InGaP transistor. The technical literature specifies performance down to DC, which is not strictly true as these amplifiers must be AC coupled at both the input and the output. The schematic of the Darlington and its

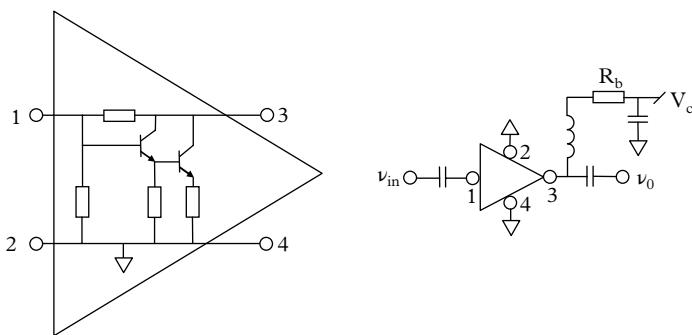


Fig. 14.23. The circuit of the four-terminal Darlington, left, and a $\times 10$ gain amplifier based on this unit; R_b serves as a bias resistor and a load. The RF choke prevents high-frequency feedback to the supply.

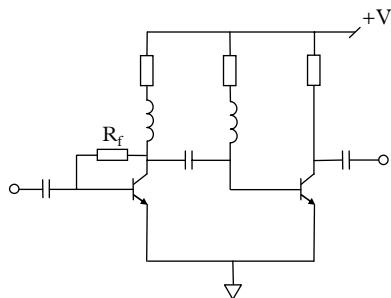


Fig. 14.24. Transimpedance amplifier described by Blanch *et al.* (1999). The transistors are of the type AT41533, producing an amplifier with a gigahertz bandwidth; R_f sets the transimpedance gain.

Blanch, O. *et al.* (1999). Performance of a fast low noise front-end preamplifier for the MAGIC imaging Cerenkov telescope. *IEEE Trans. NS.* 46, 800–5. Reprinted with permission from IEEE. ©1968.

biasing arrangement are given in an application note (Mini-Circuits 2009) available from their website. The point to note is the textbook simplicity of the circuit—a common trait in high-frequency amplifiers. A transistor of similar performance, the ADA-4743, is offered by Avago Technologies and based on silicon bipolar technology; the amplifier is configured in exactly the same way.

A fast, discrete-component amplifier described by Blanch *et al.* (1999) bears conceptual similarities to the configurations used in the commercial circuits already described. The preamplifier shown in Fig. 14.24 uses shunt feedback as described by Cherry and Hooper. It was developed by the MAGIC air-imaging Cerenkov telescope group for amplifying signals from an HPD device. Because of the high night-sky background, PMTs are operated at a gain of only 8000 but are

used in an application that calls for single-photoelectron detection. Its simple design, shown conceptually in Fig. 14.24, belies its exceptional performance. The outstanding feature is its $2\text{ k}\Omega$ transimpedance gain, producing an output pulse of 6 ns half width for single-photoelectron pulses. A second voltage-gain stage of 5 produces a single-electron peak of 5 mV amplitude.

14.11.3 Fast bench-top and NIM preamplifiers

These current-sensitive preamplifiers are available in boxed form or as NIM units. Fifty-ohm input and output impedance is a feature common to all such units, and the transfer function is simply $v_o(t) = 50 i_a(t)A$, where A is the specified dimensionless voltage gain. Because the input signal current is converted to a voltage at the input, these preamplifiers are usually referred to as voltage or fast-timing amplifiers.

Femto's 200 MHz $\times 10$ gain boxed unit, the HVA-200M-40-B voltage amplifier, offers the option of AC or DC coupling and a rise time of 1.8 ns. Noise at the output is easily determined from the input figure of $3.5\text{ nV}/\sqrt{\text{Hz}}$ quoted by the manufacturer. Allowing for the gain of $\times 10$, the rms noise at the output is $3.5 \times 10^{-9} \times (200 \times 10^6)^{1/2} \times 10 = 0.5\text{ mV}$, hardly discernible on an oscilloscope. There is a 500 MHz version in the same product range, and a 1.1 GHz AC-coupled amplifier in the HAS range; this amplifier is better suited to microchannel PMTs than to conventional PMTs. The ORTEC 9305 fast preamplifier has a similar specification to the HVA series but with less than $1\text{ }\Omega$ output impedance (pay attention to matching). The 9306, a 1 GHz amplifier, is similar to the HAS-X-1-40 and it is also AC coupled. A common feature of the ultra-fast Gigahertz amplifiers is AC coupling, with a concomitant rate effect.

SRS offers a four-channel, bench-top amplifier that is DC coupled, with 350 MHz bandwidth and a gain of 5 per channel. Rise time is 1 ns per channel, and cascading up to a gain of 625 is permitted but at some loss of bandwidth and rise time.

Fast NIM-based amplifiers with 4, 8, or 16 independent gain channels are available from a limited number of manufacturers. The N979 from CAEN provides 16 channels, each of $\times 10$ gain and with a rise time of 1.5 ns; these units may be cascaded.

14.11.4 Voltage- and current-feedback OP-AMPS

When discussing amplifiers for pulsed signals, it is helpful to classify PMT applications into two categories: slow and fast. Somewhat arbitrarily, we set the demarcation at a bandwidth, B , of 10 MHz or, equivalently, a rise time $t_r = 0.35/B = 35$ ns. Examples of slow applications may be found in scintillation spectroscopy, based on inorganic scintillators, whereas the use of plastic scintillator or the Cerenkov effect produces fast pulses of rise time less than 2 ns. In general, slow applications can be satisfied with ‘textbook’ solutions, for which an extensive range of voltage-feedback OP-AMPs (VFAs) is available. High-frequency applications

are more demanding, with the availability of only a limited range of suitable current-feedback OP-AMPS (CFAs).

The use of VFAs is covered by manufacturers' application notes, and attention will now centre on the use of current-feedback types. Although CFAs have been available for some 20 years, there is still some reluctance over their adoption. This perhaps stems from the fact that the inverting input is of low impedance, unlike the familiar VFAs, for which both inputs are of high impedance. There is a further challenge for those new to CFAs, concerning the substitution of closed loop gain, a key parameter in the voltage-feedback mode of operation, with impedance, $|z_0|$. This is not an arcane concept when it is realized that, by connecting the anode of a PMT to a load resistor, one is in effect creating a transimpedance amplifier. CFAs can be analysed in a fashion that is analogous to the way VFAs are analysed, and most circuits built with VFA amplifiers can be realized with the CFA types, but with better high-frequency response.

In a VFA, the output voltage is an amplified version of the error voltage, $v_p - v_n$. In the CFA shown in Fig. 14.25, there is a unity gain buffer between the two inputs, providing, ideally, infinite impedance at the non-inverting input, and zero

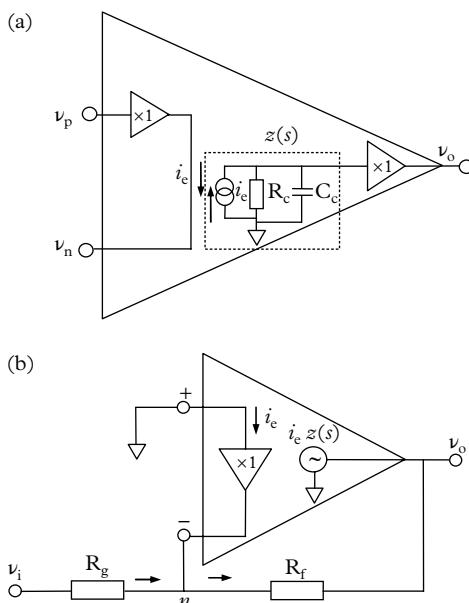


Fig. 14.25. (a) Equivalent circuit of a CFA and (b) a CFA configured for signal inversion. The expression for the closed-loop gain follows by application of the Kirchhoff current law at the node n .

Giachero, A., Gotti, C., Maino, M., and Pessina, G. (2011). Current feedback operational amplifiers as fast charge sensitive preamplifiers for photomultiplier read out. *J Inst.* 6, P05004, 1–19. Reprinted with permission from IOP ©2011.

impedance at the inverting input (in practice, less than $50\ \Omega$). The essential function is one of ‘mirroring’ the error current, i_e , to the output, with the equivalent circuit of a voltage source, v_o . CFAs have the advantage at high frequencies of achieving wide bandwidth that is substantially independent of gain and dictated by the feedback resistor R_f .

Summing the currents at node n ,

$$v_i/R_g + i_e + v_o/R_f = 0. \quad (14.18)$$

Noting that $i_e = v_o/z(s)$ and dividing by v_i leads to the expression for the closed loop gain:

$$\nu_o/\nu_i = \frac{-R_f/R_g}{1 + R_f z(s)}. \quad (14.19)$$

It is clear from (14.19) that the closed loop gain is independent of the frequency response, provided that $z(s)$ is sufficiently large, in which case $\nu_o/\nu_i \rightarrow -R_f/R_g$, which is the expression for a VFA. The closed loop gain is essentially decoupled from the frequency term in (14.19), and only R_f controls the closed loop frequency response. In a VFA, the closed loop frequency response is subject to the gain-bandwidth product, GB , but the ideal CFA does not possess a gain-bandwidth specification. Manufacturers’ data sheets indicate the minimum acceptable value for R_f , which is referred to as R_f^* ; this is greater than $500\ \Omega$ for the OPA695, for

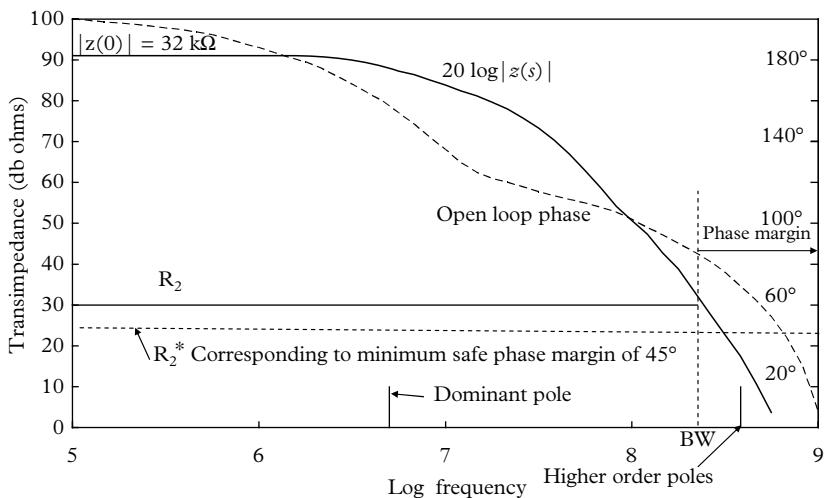


Fig. 14.26. An approximate Bode plot for the transimpedance of an OPA695; s refers to the complex variable $j\omega$.

Giachero, A., Gotti, C., Maino, M., and Pessina, G. (2011). Current feedback operational amplifiers as fast charge sensitive preamplifiers for photomultiplier read out. *J Inst.* 6, P05004, 1–19. Reprinted with permission from IOP ©2011.

example, a popular choice among OP-AMPS. Unlike its VFA counterpart, a CFA is unstable as a voltage follower, that is, when it is used in a mode in which the inverting input is connected directly to the output. A Bode plot for a CFA, of the type OPA695, is shown in Fig. 14.26.

There is useful information hidden within the curves regarding the approximate position of the poles:

- The open loop transimpedance gain at frequencies below $f_1 = 1/2\pi R_c C_c$ (the dominant pole) is $|z(s)| = |z(0)| = 91 \text{ dB} \equiv 32 \text{ k}\Omega$. In Fig. 14.26, taking $R_c = 32 \text{ k}\Omega$, and the position of the first pole at 5 MHz, as indicated in the plot, leads to $C_c \approx 1 \text{ pF}$. The curve rolls off at 20 dB per decade initially but steepens to twice this value once f passes the next pole.
- The open loop phase initially falls off in a manner dictated by the first pole, that is, as $-\tan^{-1}(f/f_1)$ and then more steeply as higher-order poles take effect. Once the phase approaches 0° , the amplifier will oscillate; standard practice is to arrange the feedback such that the phase margin never falls below 45° .

14.11.5 Transimpedance amplifier circuits using CFAs

All major semiconductor manufacturers provide application notes, and those with titles such as ‘Voltage feedback vs current feedback op amps’ (Karki 1998) and ‘Current-feedback myths debunked’ (Buck 2002) are particularly relevant to the present discussion. While these notes are essential reading for anyone faced with designing fast circuits with OP-AMPS, the manufacturers tend to provide performance figures for $\times 2$ or $\times 10$ non-inverting gain, with little information on transimpedance networks. The work reported by Giachero *et al.* (2011) is most welcome, for not only do they provide circuit analysis of CFAs, but they also present critical pulse performance data for three state-of-the-art gigahertz-bandwidth OP-AMPS, operated in the circuits shown in Fig. 14.27. Fast circuits have been constructed using OP-AMPS with gigahertz bandwidths, and slew rates of the order of volts per nanosecond. Ground-plane construction is essential, and surface-mounted components are preferred over leaded ones. Noise figures are similar to those for the more general OP-AMPs—in the region of $2 \text{ nV}/\sqrt{\text{Hz}}$.

Performance parameters for two of the CFAs are summarized in Table 14.2. Giachero *et al.* (2011) provide detailed circuit analysis of three different CFAs, backed by performance measurements. The parameters R_c and C_c for the equivalent circuit shown in Fig. 14.25(a) are not provided by manufacturers, and yet knowledge of their magnitudes, especially for C_c , is critical to completing any circuit analysis. Giachero *et al.* state that processing variations led to a spread of 30% in these parameters; this result was presumably deduced by measuring the performance of a representative sample of devices. The output generated by a step input, in the absence of feedback, takes the usual form,

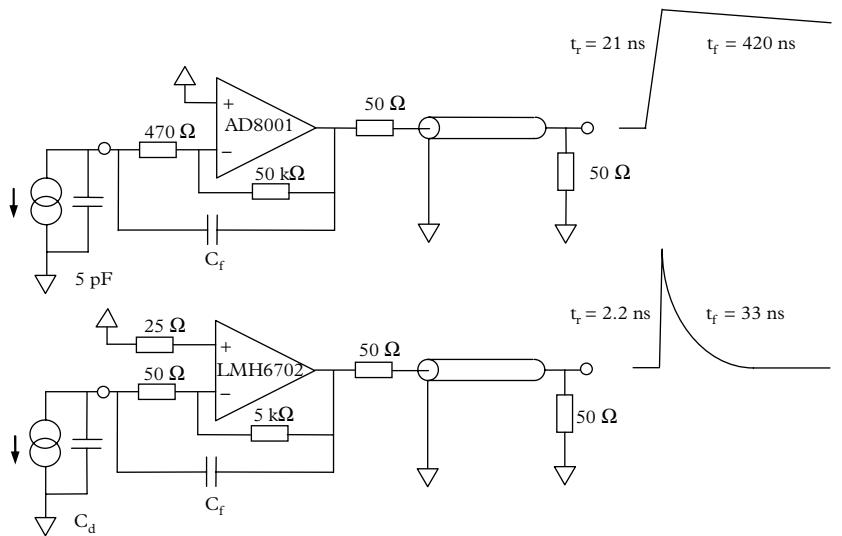


Fig. 14.27. Transimpedance amplifiers based on current feedback OP-AMPS, analysed by Giachero *et al.* (2011). Actual measurements show considerable DC offset, which can be annulled by the usual methods.

Giachero, A., Gotti, C., Maino, M., and Pessina, G. (2011). Current feedback operational amplifiers as fast charge sensitive preamplifiers for photomultiplier read out. *J Inst.* 6, P05004, 1–19. Reprinted with permission from IOP ©2011.

Table 14.2 Performance parameters for the CFAs AD8001 and LMH6702.

Parameter	AD8001	LMH6702
C_C (pF)	2	3
R_C (kΩ)	100	100
GB (GHz)	0.8	1.7
Slew rate (V/ns)	1.2	3.1
R_g (Ω)	470	50
R_f (kΩ)	50	5
C_f (pF)	1.8	5
C_D (pF)	20	20
t_r (ns)	21	2.2
t_f (ns)	420	33

$$v_o(t) = \frac{-Q}{C_C} \exp\left(\frac{-t}{R_C C_C}\right), \quad (14.20)$$

from which C_C can be deduced from Q/C_C , the measured peak output voltage; R_C is estimated from the decay time of the pulse. This procedure is illustrated by the authors for an OPA695 with a slew rate of 4.3 V/ns, the fastest of all the OP-AMPS examined. A circuit will oscillate if R_f is below a certain value, usually a few hundred ohms and sometimes quoted by the manufacturer (e.g. 237 Ω for the LMH6702). The operating bandwidth is dependent upon the signal amplitude as quantified by the slew rate. Rise time is primarily determined by the position of the high-frequency pole; it can easily double in going from small signal conditions (pulse amplitudes up to 100 mV) to large-amplitude pulses of 1 V. Output without feedback for a step input is given by

$$v_o(t) = \frac{-Q}{C_f + C_C} \exp\left(\frac{-t}{R_f(C_f + C_C)}\right). \quad (14.21)$$

The role of R_f is to reduce the fall time, and its adopted value does not affect gain or stability; C_f provides the means for reducing sensitivity. As previously mentioned, regarding charge-sensitive preamplifiers, fast amplifiers of all types are best located in a way that minimizes the length of input signal leads: that is, on the same board as the voltage divider (with the ground plane), or on a second board positioned immediately behind the first one. Matching can be relaxed with close coupling, allowing the designer maximum scope in the selection of input circuitry. For example, it is usually possible to increase the input impedance of a preamplifier, thereby enhancing the voltage gain, with limited loss of rise time. This does not apply when the preamplifier is remote from the anode, and connected via coaxial cable. The options available when coaxial cables are involved have already been discussed in 14.3.1.

14.11.6 Logarithmic amplifiers

Logarithmic axes are frequently used in science for the optimal presentation of graphical information. Either or both axes may be expressed in logarithmic terms to best represent the data. One way of improving dynamic range is to include a log amp between the PMT and an ADC. One bit of an ADC now corresponds to a fixed fraction of the signal over its range. The degree of detail in spectral distributions is thus constant over the entire energy range. Perhaps the earliest application of this technique was the detection of cosmic ray events spanning 10^9 to 10^{14} eV. Radeka (1973) made use of a bipolar transistor for the logarithmic feedback element in a fast-charge amplifier. More recently for LIDAR studies, Lienert *et al.* (2002) were able to extend the dynamic range of a PMT from two to four decades by using a fast log amp of 50 MHz bandwidth. Femto offer a boxed log amp with a 60 dB dynamic range and a 5 ns rise time, while the AD8307 is commercially available with a bandwidth of 500 MHz and 92 dB dynamic range.

Achieving dynamic ranges of up to 100 dB is subject to linear PMT performance. Early versions of log amps were found to be temperature sensitive and subject to gain drift; such shortcomings appear to have been overcome. However, given their apparent advantages in covering a wide dynamic range, the reasons for their limited adoption are unknown.

14.12 Pulsed-light sources

PMT users wishing to calibrate or optimize timing performance have found the avalanche transistor-based sub-nanosecond light pulser particularly useful. Two centres of excellence for timing measurements were established for this purpose: one by Leskovar and Lo in the USA, and one by Moszyński in Europe. Manufacturers since the 1970s have found it beneficial to submit new products to these organizations for critical evaluation. Instrumentation based on Ferranti XP22 and XP23 gallium arsenide LEDs offers the advantage of the fastest switching attainable, but with the disadvantage of emission in the red part of the spectrum, and relatively high cost. Ideally, as discussed in 8.1.2, calibration should always be carried out at the wavelength of intended application. The availability since 2003 of LEDs with peak emission in the UV and blue regions of the spectrum allowed calibration at scintillator and Cerenkov wavelengths, for example, the Nichia NSHU590E at 370 nm, and the NSPB300 at 370 nm. The temperature coefficient of LEDs is known to be high—Vićić *et al.* (2003) quote a figure of 2.2%/°C for the NSHU550E LED—and so temperature stabilization may be necessary. A particular large-scale requirement stems from massive astroparticle physics detectors, which contain thousands of PMTs, with ideally one light pulser per PMT. This has cost implications which have long since been met by the choice of inexpensive components.

Passive pulsed-light sources can be devised using LEDs, although the equivalent circuit is complex—characteristically, a few hundred picofarads in series with a resistance of $\sim 10 \Omega$ (Vićić *et al.* 2003). Reasonably fast light pulses (~ 10 ns width) can be generated, given a drive pulse from a fast laboratory pulse generator. This is illustrated in Fig. 14.28(a) for coupling to 50Ω coaxial cable. Some scope for selecting R_T for optimal pulse shape is achievable. However, significantly faster light pulses can be derived by using an avalanche circuit, shown in its rudimentary form in Fig. 14.28(b), where the generation of the drive pulse for the LED is illustrated.

Laser-diode-based picosecond light pulser (PLP) instruments offered by Hamamatsu provide pulses of 50 ps width, operating at a rate of up to 100 MHz. There are ten options in the series, spanning the range 375 to 1550 nm, but the PLP10-38 and PLP10-40, operating at 375 nm and 405 nm, respectively, are of particular interest to the PMT user. The driver circuits bear similarity to the aforementioned APD versions, with a laser diode and a photodiode included in the same package.

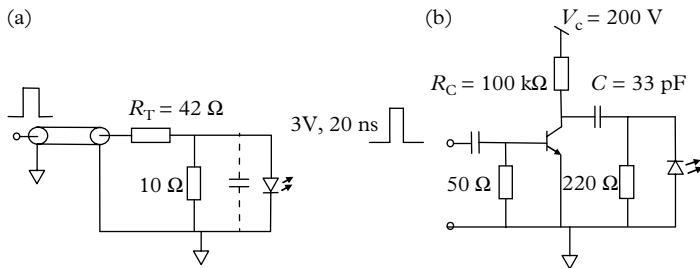


Fig. 14.28. (a) An elementary circuit for producing light pulses of 10 ns characteristic width. (b) An avalanche transistor circuit capable of producing sub-nanosecond-wide pulses from a variety of LEDs.

14.13 Optimizing PMT performance

The basic mode of PMT operation may be achieved by connecting the anode to a load resistor, R_L , either directly or via a capacitor. This arrangement performs conversion from a current generator to a voltage source, and R_L can be chosen to match coaxial cable if required. Choosing R_L with a resistance greater than 50Ω enhances voltage gain, as is evident in (14.6), but with an increase in rise time of the leading edge, which may not be important. The use of external electronics modules to supplement PMT gain may be necessary to avoid performance limitations when PMTs are operated in this way. Such circuits, or modules, may also provide active filtering and pulse shaping. Each sub-system included in a complete detector has a range over which satisfactory performance may be obtained. For example, a given application may be fulfilled with a PMT gain set between 10^5 and 10^6 . Operating at the higher end of the PMT gain scale would undoubtedly reduce the performance demanded of any amplifier that may be required in the system, and vice versa. It is important to recognize that PMTs have a finite operating range, in particular with respect to non-linearity in gain and reduced lifetime. Both these effects are mitigated by a reduction in PMT gain.

PMTs have a half-life of the order of 200 coulombs, that is, the multiplier gain drops by a factor of 2 after this charge has been drawn from the anode, but photocathode sensitivity is generally unaffected. To put this in context, a PMT operating continuously and drawing an initial anode current of $10 \mu\text{A}$ generates ~ 300 coulombs/year at the anode. Clearly, it would be sensible to reduce the PMT gain by a factor of 10 or even 100, compensating this by external amplification. The loss in gain relates to the mean anode current, and whether it is DC or pulsed is irrelevant. The appropriate amplifier would be one of narrow bandwidth for DC applications, and one of broader frequency response for fast signal pulses.

It has already been shown in (13.21) and (13.22) that voltage dividers comprising solely resistors cannot provide absolutely stable biasing. As a consequence,

PMT gain varies in the presence of changing mean anode current. The degree of non-linear amplification in a PMT is related to the standing current in the voltage divider, I_{D0} . A well-established empirical rule states that I_{D0} should be 100 times greater than the maximum anticipated DC anode current, in which case, the gain change will be restricted to less than 1% for I_a ranging from zero up to the maximum design value. In practice, to achieve this performance, $I_a(\text{max})$ must be less than $\sim 10 \mu\text{A}$, which implies a standing current, I_{D0} , of 1 mA. Apart from unwelcome heating effects, caused by power dissipation in the divider, there is a limited choice of commercial power supplies capable of providing this level of current. Active dividers, discussed in 13.6, can maintain stable interdynode voltages for anode currents approaching 300 μA , a current level beyond the capability of most PMTs, and certainly outside manufacturers' specifications. In addition to gain changes induced by unstable biasing, there are limitations in PMT performance itself. Certain high-gain box-and-grid PMTs are only capable of maintaining stable gain for anode currents approaching 10 μA —this is attributed to space charge effects.

PMTs are limited in their ability to amplify linearly with respect to total charge and pulse height. The upper limit on pulse height extends to about 150 mA, depending on the type of multiplier and the details of the divider design. Note that 100 mA developed across 50 Ω corresponds to a peak signal of 5 V, which is close to the top end of the dynamic range offered by commercial amplifiers. Linearity performance figures given by manufacturers generally refer to the peak current at which there is a deviation, of less than a stated percentage, from linear amplification: most specifications are positioned between 1 and 5%. Reducing the PMT gain by a factor of 10 and compensating for this through the addition of a voltage amplifier of gain 10 leaves the system sensitivity unchanged, but the reallocation of gain results in a marked improvement in linearity. An amplifier of quality is needed if all the PMT attributes are to be retained: it must be stable, with low noise, a wide bandwidth, and a high slew rate, as well as being low cost.

14.14 DSP

In DSP the anode current waveform is sampled by a flash ADC, thereby representing a pulse by a sequence of numbers $h(t_0), h(t_1), \dots, h(t_n)$; Q_a , the total charge in an event, is proportional to the sum of these numbers. Resolution is determined by the number of bits and by the sampling rate, which can be as fast as 5 GS/s with 14-bit precision (i.e. a reading every 0.2 ns). The power of DSP lies in being able to extract the maximum information from a PMT. Shaping processes that were previously the preserve of analogue electronics can now be done by digital filters. Where timing is concerned: threshold, zero crossing, and CF methods can all be realized on-line.

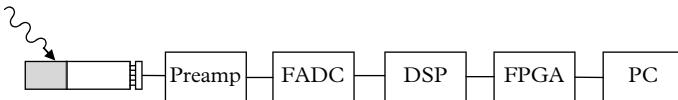


Fig. 14.29. A DSP arrangement for gamma ray spectroscopy. The terminology is explained in the text; PC, personal computer.

In analogue systems analogue-to-digital conversion is the last executed process. The reverse applies in DSP, in which conversion is performed immediately after the detector, or, after a preamplifier. It is clear from the narrow sample times involved that data output rates will be extremely high, requiring a field-programmable gate array (FPGA) to perform on-line digital pulse processing. An FPGA has the capability of performing analogue processes, such as pulse shaping, pulse height analysis, CF timing, baseline restoration, single-channel analysis, and PSD, by digital means. In fact, all operations that can be done in the analogue domain can be performed by DSP. Only the essential information derived from a sequence of digitized readings is passed on to a computer by the FPGA, for further analysis and display. A typical configuration is shown in Fig. 14.29.

The output from the flash ADC (FADC) is a series of digital samples continuously read by a FPGA and stored in a memory buffer. On command of a trigger, the contents of the buffer are frozen and data acquisition continues in a second buffer. The stored information allows interrogation of the data stream, moving backwards in time. Data recorded before the trigger can be recovered and used to establish the baseline just prior to the event—the effects of a moving baseline on rate are thus eliminated. Stored information can also serve in the search for small-sized prepulses, as shown in Fig. 14.30(c).

The classical method for extracting energy or timing information from a current pulse is covered in 14.9.1 and 14.11. This is the analogue approach, in which we seek a number to quantify the size of the pulse (i.e. its total charge, Q_a) and another to represent the time at which the event took place (the time stamp). The task is eased when only one of these quantities is required. Otherwise, the PMT signal has to be split in the manner illustrated in Fig. 14.8. The essential difference between the analogue and the digital approach is illustrated as follows. In the analogue method, the current pulse is shaped in a two-step process: first, by a charge-sensitive preamplifier and then by a shaping amplifier that produces a near-Gaussian shaped pulse, the height of which is proportional to the total charge, Q_a , in the original signal. The height, h , is then converted to a number by a peak-sensing ADC. In the digital approach, pulse shaping is done by an on-line trapezoidal filter, illustrated in Fig. 14.31. Whereas the shaped pulse from a shaping amplifier has a rounded top occurring at the set peaking time, the pulse shaped by a trapezoidal filter has better fidelity because of its flat top.

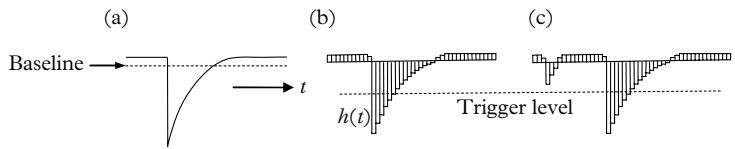


Fig. 14.30. (a) Digitization of an analogue signal waveform. The effect of baseline shift is clearly visible in (b). The precursor pulse in (c) is unrecorded in analogue systems.

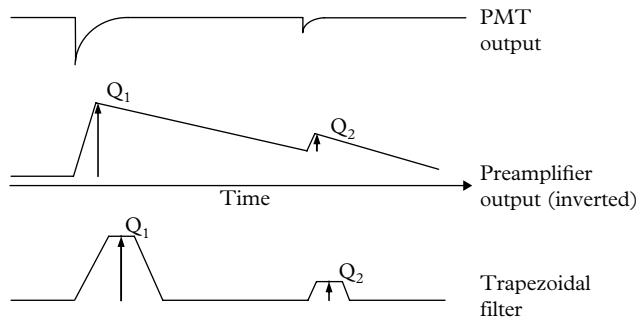


Fig. 14.31. An energy-encoding system based on trapezoidal filtering. The filter performs the function of a shaping amplifier; Q_1 and Q_2 represent the charge content of each preamplifier pulse.

A useful comparison between an analogue system, based on a charge ADC (QDC), that integrates each PMT pulse directly, and a wholly digital pulse processing arrangement is discussed by CAEN SpA (2010) in an application note available from their website. A quick comparison shows that the digital system produces a resolution of 7% with NaI(Tl) and ^{137}Cs while, for a traditional analogue system, the comparable figure is 8.2%.

A second application, worthy of discussion, is the use of PSD for particle identification among electrons, protons, neutrons, ..., and ^4He . The method is based on the observation that certain scintillators respond in a way that is peculiar to the type of incident radiation: the time signature of the scintillation is different for each particle type. The difference is small and lies in the pulse shape, with neutrons and alphas producing relatively more light output in the tail of the event than do other particles. There is a range of techniques that may be used for discrimination, among which are those based on a shaped pulse and zero crossing, or, comparison of the total signal, Q_L , and a gated version, Q_S , centred on the main pulse. There are other techniques and a considerable literature on the traditional methods of PSD, which is described in Nicholson (1974). The digital technique is explored in an application note by CAEN (2011), AN2506, which is based on evaluating the fraction $(Q_L - Q_S)/Q_L$.

REFERENCES

- Baturitsky, M. A., and Dvornikov, O. V. (1999). Advanced low input resistance preamplifiers for short input current pulses. *Nucl. Instr. and Meth. in Phys. Res. A*, **432**, 163–73.
- Blanch, O. *et al.* (1999). Performance of a fast low noise front-end preamplifier for the MAGIC imaging Cerenkov telescope. *IEEE Trans. NS*, **46**, 800–5.
- Buck, A. (2002). *Current-Feedback Myths Debunked*. National Semiconductor Publication OA-20. National Semiconductor, Arlington, TX. Available at <http://cas.ee.ic.ac.uk/people/dario/files/E416/OA-20.pdf> (accessed 7 Feb 2017).
- Busso, L. *et al.* (1999). A fast preamplifier for proximity focussed HPDs. *Nucl. Instr. and Meth. in Phys. Res. A*, **432**, 436–9.
- CAEN. (2010). *Application Note AN2503. Charge Integration: Analog Vs. Digital*. CAEN SpA, Viareggio. Available at http://www.caen.it/documents/News/26/AN2503_Charge_Integration_analog_digital.pdf (accessed 7 Feb 2017).
- CAEN. (2011). *Application Note AN2506. Digital Gamma Neutron Discrimination with Liquid Scintillators*. CAEN SpA, Viareggio. Available at http://www.caen.it/documents/News/26/AN2506_Digital_GN_discrimination_01.pdf (accessed 7 Feb 2017).
- Cherry, E. M. and Hooper, D. E. (1963). The design of wide-band transistor feedback amplifiers. *Proc. IEE*, **110**, 375–89.
- Collinge, B., West, C., and Lloyd, G. H. (1965). A high stability amplifier for nuclear physics applications. *Nucl. Instr. and Meth.*, **35**, 313–16.
- de Vries, J. and Kelling, F. E. T. (1988). Fast timing with photomultiplier dynode pulses. *IEEE Trans NS*, **35**, 392–5.
- DIRC. (1997). *Grounding and Shielding*. DIRC Electronics Final Design Report. Laboratoire de Physique Nucléaire et de Hautes Energies, Paris.
- Ernwein, R. and Gorodetzky, Ph. (1976). Green light emitting diodes simulate nuclear events. *Nucl. Instr. and Meth.*, **138**, 57–9.
- Franzen, D. L. and Day, G. W. (1979). LED source for determining optical detector time response at 1.06 μm. *Rev. Sci. Instr.*, **50**, 1029–31.
- Giachero, A. *et al.* (2011). Current feedback operational amplifiers as fast charge sensitive preamplifiers for photomultiplier read out. *J. Instr.*, **6**, 1–19.
- Gillespie, A. B. (1970). ‘Low-noise head amplifiers’. In *Electronics for Nuclear Particle Analysis*, ed. L. J. Herbst, Oxford University Press, Oxford, 138–60.
- Goyot, M. (2000). A versatile fast low noise preamplifier in UHF bipolar technology. *Nucl. Instr. and Meth. in Phys. Res. A*, **442**, 374–7.
- Herbst, L. J. (1970). *Electronics for Nuclear Particle Analysis*. Oxford University Press, Oxford.
- Hussain, K. A. and Butt, D. K. (1980). Time resolution studies of fast photomultipliers. *Nucl. Instr. and Meth. in Phys. Res. A*, **173**, 471–5.

- Kapustinsky J. S. *et al.* (1985). A fast timing light pulser for scintillation studies. *Nucl. Instr. and Meth. in Phys. Res. A*, **241**, 612–13.
- Karki, J. (1998). *Voltage Feedback vs Current Feedback Op Amps*. Texas Instruments, Dallas, TX. Available at <http://www.ti.com/lit/an/slva051/slva051.pdf> (accessed 7 Feb 2017).
- Keithley Instruments Inc. (1972). *Low Level Measurements Handbook* (7th ed.). Tektronix, Beaverton, OR. Available at http://www.tek.com/sites/tek.com/files/media/document/resources/LowLevelHandbook_7Ed.pdf (accessed 29 Jan 2017).
- Knoll, G. F. (1989). *Radiation Detection and Measurement* (2nd ed.). John Wiley and Sons, Inc., New York.
- Landis, D. A., Madden N.W., and Goulding, F. S. (1998). Bipolar pulsed reset for AC coupled charge sensitive preamplifiers. *IEEE Trans. NS*, **45**, 805–9.
- Lienert, B. *et al.* (2002). A 50 MHz logarithmic amplifier for use in Lidar measurements. *J Atmos. Oceanic Technol.*, **19**, 654–57.
- Lubsandorzhiev, B. K. and Vyatchin, Y. E. (2006). Studies of Kapustinsky's light pulser timing characteristics. *J. Instr.*, **1**, T06001.
- Lunsford, J. S. (1964). Nanosecond pulse amplifiers. *Rev. Sci. Instrum.* **35**, 1483–5.
- Madden, N. W. *et al.* (1969). An improved operating mode for a Si(Li) x-ray spectrometer. *IEEE Trans. NS*, **37**, 171–6.
- McFarlane, W. K. (1973). An inexpensive nanosecond light pulser for use in photomultiplier testing. *Rev. Sci. Instrum.* **45**, 286–9.
- Millard, J. K. (1967). New shunt-series nanosecond pulse amplifier. *Rev. Sci. Instr.*, **38**, 169–75.
- Millman, J. and Halkias, C. C. (1972). *Integrated Electronics*. McGraw-Hill, New York.
- Mini-Circuits (2009). *Biasing MMIC Amplifiers* (e.g., ERA Series). Application note AN-60-010. Mini-Circuits, Brooklyn, NY. Available at <https://www.minicircuits.com/pages/pdfs/an60010.pdf> (accessed 7 Feb 2017).
- Mori, K. (1996). Charge sensitive amplifier, radiation detectors and their uses. *Proc. 10th Workshop on Radiation Detection and Their Uses*. KEK, Japan.
- Nicholson, P. W. (1974). *Nuclear Electronics*. John Wiley and Sons, Inc., London.
- Nowlin, C. H. and Blankenship, J. L. (1965). Elimination of undesirable undershoot in operation and testing of nuclear pulse amplifiers. *Rev. Sci. Instr.*, **36**, 1830–40.
- Ott, H. W. (1988). *Noise Reduction Techniques in Electronic Systems* (2nd ed.). John Wiley and Sons, Inc., New York.
- Pasquali, G., *et al.* (2007). A DSP equipped digitizer for on-line analysis of nuclear detector signals. *Nucl. Instr. and Meth. in Phys. Res. A*, **570**, 126–32.
- Radeka, V. (1970). Charge amplification without charge leak resistor. *IEEE Trans NS*, **17**, 433–9.
- Radeka, V. (1973). Logarithmic charge amplifier for scintillation detectors. *Nucl. Instr. and Meth.*, **113**, 401–12.
- Radeka, V. (1988). Low-noise techniques in detectors, *Ann. Rev. Nucl. Part. Sci.* **38**, 217–77.

- Rush, C. J. (1964). New technique for designing fast rise transistor pulse amplifiers. *Rev. Sci. Instr.* **35**, 149–56.
- Shubnikov, E. I. and Subbotin, F. M. (1972). On the shape of the one electron pulse from a photoelectron multiplier. Translated from *Pribory i Tekhnika Eksperimenta*. No. 1, 1973, Plenum Publishing Corporation, New York.
- Spieler, H. (2001). *Radiation Detectors and Signal Processing. Lecture Notes VII. Heidelberg Graduate Lectures in Physics. University of Heidelberg Oct. 8–12, 2001*. Available at http://www-physics.lbl.gov/~spieler/Heidelberg_Notes/ (accessed 7 Feb 2017).
- Vićić, M. et al. (2003). Fast pulsed UV light source and calibration of non-linear photomultiplier response. *Nucl. Instr. and Meth. in Phys. Res. A*, **507**, i.

Appendix A

PMT output distributions

A.1 General considerations

Different authors plot different types of graphs for PMT output distributions. Integral and differential representations are possible and the abscissa may be either pulse height, h , or HV, V . We are interested in two things: the connection between these representations, and which one is most appropriate for a given application. In mathematical probability theory, it has been found formally convenient to use integral distributions as the fundamental ones, $P(>h)$, for a continuous variable h . One distinguishes this from the pdf, $p(h)dh$, which is a differential distribution, and means the fraction of events between h and $h + dh$. The distributions are normalized if

$$P(>0) = 1, \text{ and } \int_0^\infty p(h)dh = 1, \quad (\text{A.1})$$

and the two functions are always related by

$$p(h) = \frac{dP}{dh}(>h). \quad (\text{A.2})$$

Note that $P(h)$ and hence also $p(h)$ may be continuous or discrete, depending on the nature of h . Signals at the photocathode are always discrete, whereas at the anode signals are, for practical purposes, continuous. We can say this because even a signal initiated by a single photoelectron will result in $\sim 10^6$ or more electrons at the anode, and the time signature of the event is therefore essentially continuous. The statistics that apply to the detection of the light signal, at the photocathode, are always discrete (usually binomial) but the distribution at the anode is effectively a continuous one and, as will become evident, the distribution generated by the available signal handling electronics is discrete. In this appendix, we are only concerned with distributions at the anode.

Signals at the output of a PMT consist of short-lived charge pulses, each of which correlates with the detection of a single photon at the photocathode. The signals that concern us are those that are sufficiently separated in time so that they seldom overlap, and may therefore be classed as independent. For example, a feeble DC light source, producing, say, 1000 pe/s, certainly satisfies this requirement. Signals containing more than one photoelectron may also satisfy the independence condition, provided every pulse of photons occurs over a short time interval, τ .

Specifically, if $\tau \ll 1/R$, where R is the average rate of events, then these multi-photon events will be individually resolved. We obtain the maximum possible information about a light source by sizing and counting each event at the anode (the actual time of arrival of each signal is irrelevant for present purposes). Signals generated in high-light-level applications, where $R \geq 1/\tau$, are generally handled by DC techniques. In highly specialized applications, the signals may be digitized at a very fast rate for subsequent software processing, with the capability of separating overlapping events. We will consider the four different ways for presenting the same set of data, together with schematics of the instrumentation applicable in each case. The type and means of data collection depend on the application: in some cases, the requirement is to simultaneously determine the rate of the events and the individual size of each event; in others, only the rate is required.

An MCA is an instrument that is capable of simultaneously counting and sizing signals. The memory of a typical MCA has 1024 registers, each of which records the number of events within a narrow size band, and a representation of the spectrum of the light source is obtained by recording and analysing a large number of events. Since the channel width of the instrument is fixed, the effect of increasing the PMT gain is to spread the number of events recorded over more channels (see A.7, Fig. A.5). An MCA is a type of signal digitizer in that it assigns a number to each pulse analysed. (Note, however, that the term signal digitizer is usually reserved for the process in which the output signal itself is continuously sampled at many short and contiguous intervals for the purpose of recording its profile.) The data from an MCA can be manipulated and presented in various ways to suit the experimentalist. Until quite recently, MCAs were essentially laboratory instruments but nowadays most of the portable, battery-operated radiation monitors, in the form of a handheld instrument, include an MCA.

Whatever instrumentation is used, there is always the requirement to present data in the form of integral or differential distributions—which is now our purpose. It will be shown that, given one of four distributions, it is possible to generate all of the other three—a very important facility. The differential pulse height distribution at the output of a PMT is a function of pulse size h , and of the applied HV V . Note the italic form is used to emphasize a variable quantity. The actual shape of the distribution depends on the spectrum of the source of signals, and on certain characteristics of the PMT. The distributions of particular interest are those of single-photon emission, together with those generated by organic and inorganic scintillators. However, the theory developed here is general, and applicable to any theoretical or experimentally determined set of data.

It is important to be clear on the meaning of pulse height, h . In practice, h refers to the size of the output, as shown on the abscissa of every pulse height distribution. Since a PMT generates a charge in response to a light signal, the natural unit for h is picocoulombs. However, in many practical situations the output charge is electronically processed to smooth and shape the signal so that the peak height attained by the signal is proportional to the charge (this is the origin of the term ‘pulse height’). The abscissa in this case has the units of volts or millivolts. In the context of the

inter-relationship between the distributions, we need not be specific as to the precise units of pulse height—it is sufficient to regard h as the variable on the x-axis.

Of the four pulse height distributions relevant to PMT applications, two are integral, and two are differential. Synonymous terms are cumulative and density, respectively. Integral counts refer to the total number of counts per unit time (usually seconds) that exceed a specified threshold h . The differential count rate is the number of counts with amplitude between h and $h + \Delta h$, usually expressed as counts per second per channel. The following representations provide pictorial explanations of the manner in which these distributions may be measured, their general shape, and how they interrelate.

A.2 Schematics and methods for measuring distributions

Analogue signals may be measured in various ways for later display. The source of light might be single photon, or a spectrum accumulated from a scintillator, for example. It is essential that the light source remains constant while the various measurement schemes are investigated. A family of curves may be generated for each set of variables by changing a parameter, such as the HV (see A.7, Fig. A.5).

A.3 Integral pulse height distributions

An integral pulse height distribution is designated $N(>h)|_{V_0}$, which means the number of counts greater than h in amplitude measured at fixed HV, V_0 . The entity, h , may be expressed in various units, depending on the details of the measuring system. Typically, h might be expressed as charge, millivolts, keV, or gain. An integral pulse height spectrum is also known as a cumulative count-rate distribution, measured in units of counts per unit time. Figure A.1(a) shows a versatile arrangement for measuring $N(>h)|_{V_0}$ under photon-counting conditions. Based on a 50Ω NIM system, the amplifier gain needs to be of the order of $\times 100$ for a discriminator of 50 mV sensitivity, giving an effective minimum threshold of 0.5 mV. Either the amplifier gain or the discriminator threshold may be varied to achieve the required cumulative spectrum. Alternatively, a 50Ω stepped attenuator may be incorporated either before or after the amplifier for this purpose.

A.4 Differential pulse height distributions

The differential pulse height distribution is $dN(>h)/dh|_{V_0}$, with V_0 a parameter, and h variable. The output counts are recorded as h is varied from its starting point to a terminal value that encompasses the entire spectrum. For practical purposes, the differential distribution may also be derived from the results of Fig. A.1 by subtracting adjacent readings spaced Δh apart. Clearly, a pulse height distribution

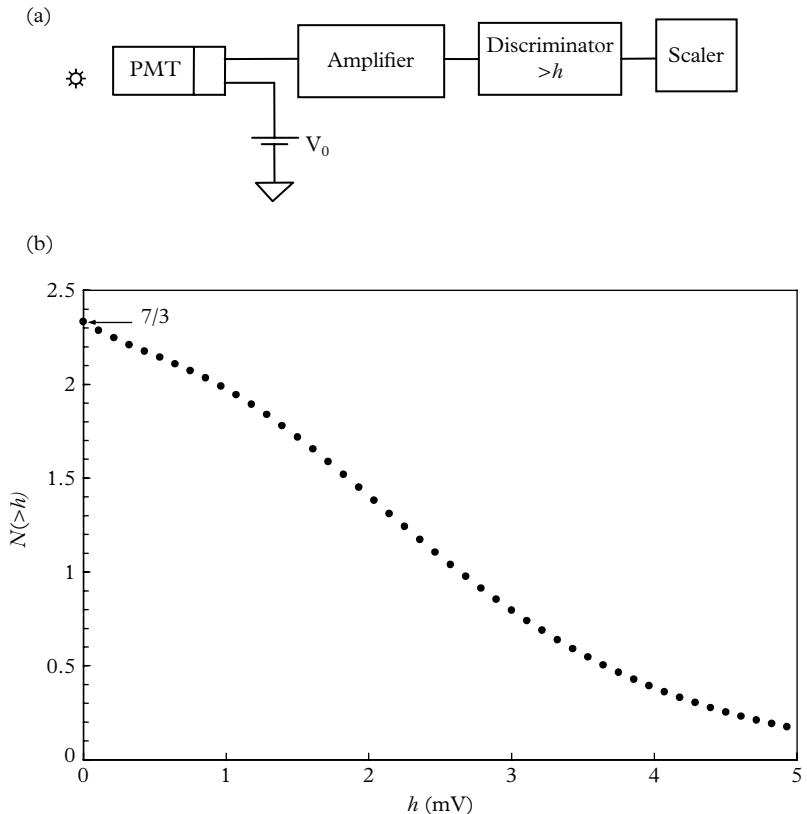


Fig. A.1. (a) A set-up for measuring integral pulse height distributions. Note for Curve (b) that the slope is always negative for this type of spectrum. The detection threshold, h , is variable, and the HV, V_0 , is fixed.

is more efficiently gathered by using an MCA in preference to scanning h . The units are typically counts per unit time per millivolt.

A.5 Plateau characteristic

The integral distribution, $N(V)|_{h_0}$, for fixed h_0 and variable V , is a function representing the contribution from all pulses in the spectrum that exceed h_0 in height. This is generally referred to as a ‘plateau’ characteristic, in view of its apparent flatness at sufficiently high V (Fig. A.3).

The differential pulse height distribution, $dN(>h_0, V)/dV|$, with h_0 fixed and V variable, is written as $n(V)|_{h_0}$, with the subscript h_0 emphasizing that this distribution is defined for all $h > h_0$ and is a function of V only. Figure A.4 shows the

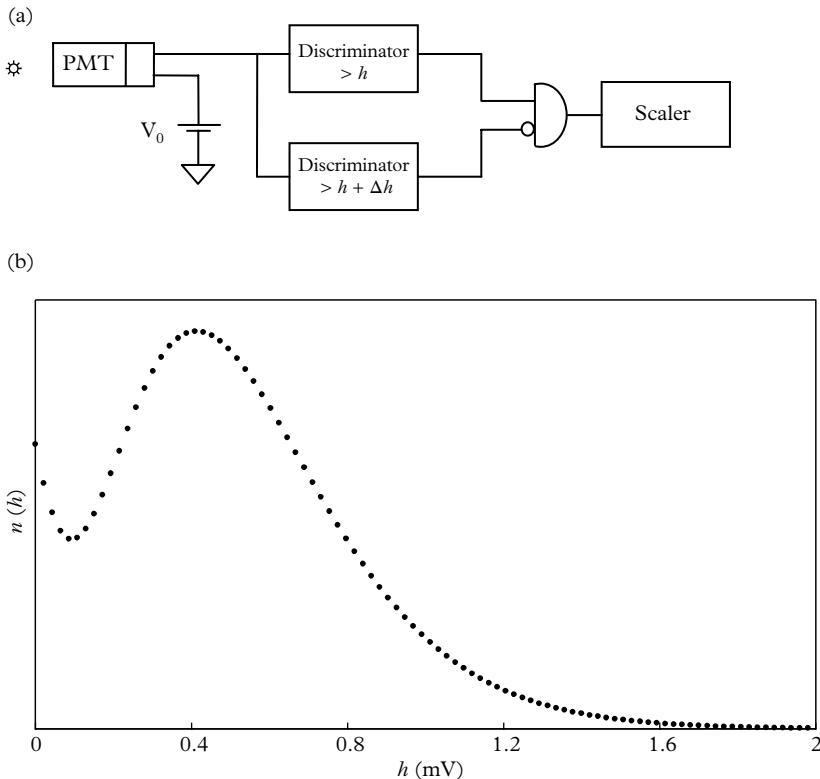


Fig. A.2. The probability density function (pdf). Also known as a differential pulse height distribution; V_0 is fixed while h is variable.

differential distribution of Fig. A.3. There is no experimental arrangement for obtaining this spectrum and it has to be deduced from the other distributions. For example, it may be derived from Fig. A.3 by subtracting adjacent readings, spaced ΔV apart. The units are typically counts per unit time per volt. Even though it cannot be measured directly, it serves as a guide for deciding upon the operating point on a given plateau characteristic: $n(V)|_{h_0}$ is the slope at any point on $N(>h_0, V)$. The use of this distribution in photon counting is particularly helpful for setting up (see A.7, and A.9, Fig. A.14).

A.6 Changing variables

We define the set of functions, illustrated in the previous sections, that will allow us to present the PMT output pulse size distribution in various ways. There are two independent variables to consider: h , the pulse height, and, V , the HV;

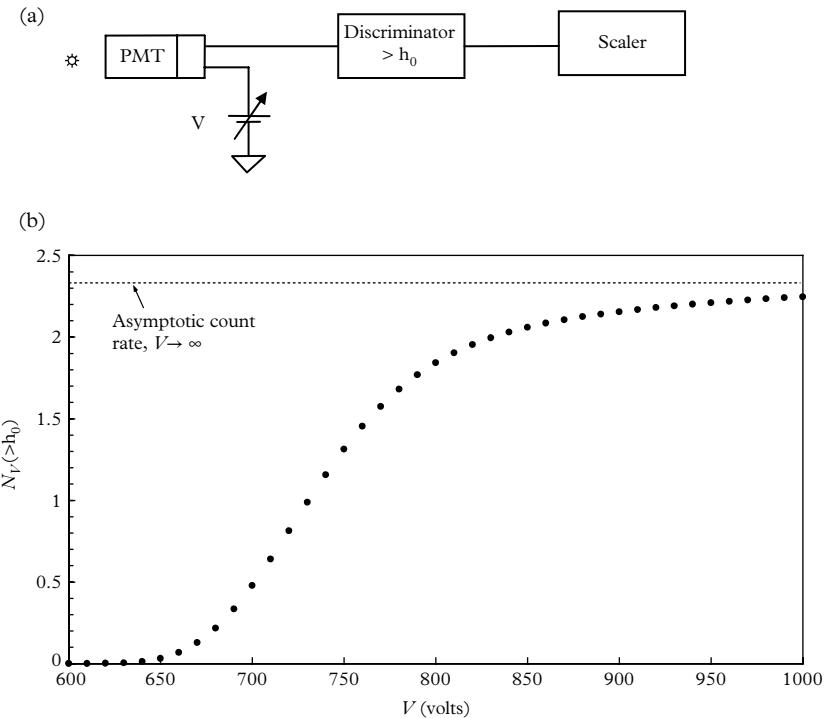


Fig. A.3. The cumulative distribution is generally referred to as a plateau characteristic. The threshold, h_0 , is fixed while V is varied. The significance of the asymptotic count rate is discussed in A.7 and in A.11.

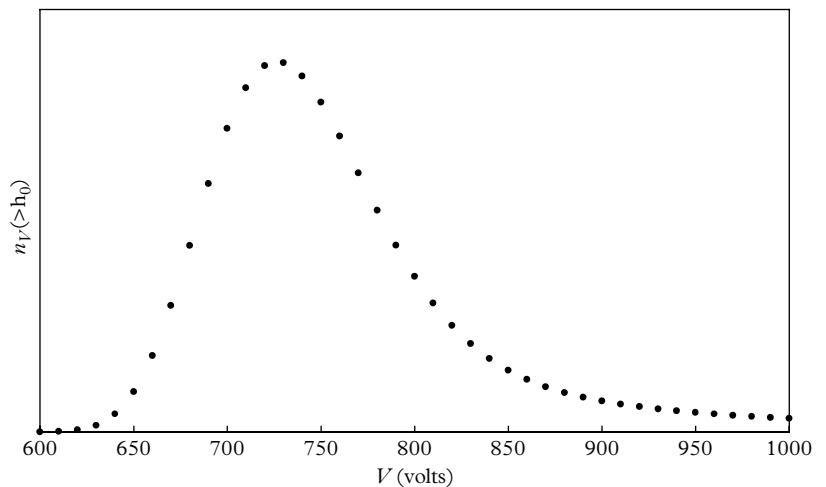


Fig. A.4. The differential distribution of the plateau characteristic $N_V(>h_0)$ deduced by subtracting adjacent readings in Fig. A.3.(b).

$N(>h, V)$ represents the integral, or cumulative, counts per unit time interval for all signals $>h$, and over all operating voltages V . This three-dimensional function is not always of practical use, but the subsets $N(>h, V_0)$ and $N(>h_0, V)$, previously discussed, represent measurable quantities that are of practical significance.

The method for transforming a distribution to another variable is treated in textbooks covering statistics or partial differentiation. We use the following equality as the starting point. Quite generally, we want to transform as follows:

$$n(h)|_{V_0} dh = n(V)|_{h_0} dV. \quad (\text{A.3})$$

If h is a known function of V , then, given the pulse height distribution $n(h)|_{V_0}$, we can derive $n(V)|_{h_0}$ from (A.3), as follows:

$$n(V)|_{h_0} = n(h)|_{V_0} |dh/dV|. \quad (\text{A.4})$$

A straightforward rearrangement of (A.4) allows us to re-express $n(h)|_{V_0}$ in terms of $n(V)|_{h_0}$. The explanation for the use of the modulus of dh/dV is found in good textbooks on statistics. A consequence of increasing V is to spread a spectrum over a wider range of the abscissa. The shift in the peak of the pdf to the right is most noticeable; as is the reduction in counts in every channel. However, since the light source is constant in output, so is the integral count rate. This behaviour is in accord with the power law of the form

$$h = h_0(V_0/V)^\beta, \text{ or } V = V_0(h_0/h)^{1/\beta}, \quad (\text{A.5})$$

which links h and V , and it follows that

$$\frac{dh}{dV} = \frac{-\beta h}{V_0} (h/h_0)^{1/\beta}. \quad (\text{A.6})$$

Given a pulse height distribution, $n(h)|_{V_0}$, measured with an MCA, for example, we can derive a pulse height distribution in terms of V , as

$$n(V)|_{h_0} = n(h)|_{V_0} \frac{\beta h}{V_0} (h/h_0)^{1/\beta}, \quad (\text{A.7})$$

where, from (A.5), $n(V)|_{h_0}$ values must be plotted at $V = V_0(h_0/h)^{1/\beta}$. Given any one of the four distributions, we can derive the other three, provided that we know $h(V)$. The relationship (A.5) between corresponding values of h and V is an inverse-power one and it is readily seen that high HVs correspond to small pulses, and vice versa. In practice, there is an important caveat underlying (A.3) to (A.7) that $n(h)|_{V_0}$ does not change its shape when different values for V_0 are chosen. We know that the SER peak-to-valley ratio tends to improve with increasing V_{k-d1} , as does the collection efficiency. In an all-resistor voltage divider, V_{k-d1} is proportional to V , and the transformations will apply over a limited range of V . Fixing V_{k-d1} with zener diodes largely overcomes these two dependencies. Knowledge of the absolute gain is unnecessary as transformations can always be made in terms of relative gain.

A.7 Transformations of an analytic function

The theory presented in A.6 is a minefield for the uninitiated, who may be attempting for the first time to transform an experimental distribution from one variable to another. For illustrative purposes, it is enlightening to choose an analytic function for $n(h)|_{V_0}$ that is continuous, easily integrated, and always finite. The differential spectrum $n(h)$ and its integral, $N(>h)$, are the assumed distributions and are given in (A.8) and (A.9), respectively:

$$n(h)dh = \exp(-3h) [(1 + (3h)^3]dh \quad (\text{A.8})$$

and

$$N(>h) = \int_h^{\infty} n(h)dh = 1/3 \exp(-3h)[(3h)^3 + 3(3h)^2 + 6(3h) + 7]. \quad (\text{A.9})$$

The function in (A.8) has the general shape of an SER spectrum, and the form of the distributions that follow from manipulating (A.8) and (A.9) will be recognized by those with practical experience in photon counting. In A.8 we explore experimental results presented in tabular form. The mathematical manipulations are the same for experimental data but are easier to follow, given the familiarity with the analytical approach. Typically, $\langle g \rangle$ will be of the order of 10^6 , but we do not need to know the absolute gain but only how it changes with V : that is, we only need to know β , which will be taken as 8, in the example that follows. A set of distributions is plotted in Fig. A.5 at fixed intervals, $dh = 0.05$, the first of which is annotated $V_0 = 1000$ volts ($1.0 \times \langle g \rangle$). The second distribution is at 950 volts, by adopting the gain scaling factor $(950/1000)^8$, which corresponds to $\langle g \rangle = 0.66 \times 10^6$, and similarly for distributions at other operating voltages. Note that the channel width is the same for all curves (0.05). Figure A.5 describes the general shape the distributions would exhibit on an MCA screen when viewing a single-photon source of fixed intensity.

Cumulative distributions, $N(>h, V_0)$, are derived from (A.9) and plotted in Fig. A.6. We note at $h = 0$ that all curves predict the same cumulative count of $7/3$, for the reasons already stated. These curves may also be obtained by integrating (A.8) numerically, as follows. For example, we take the curve annotated 850 volts in Fig. A.5 and start at an h value for which the differential count is negligible, say, at $h = 1.25$. We accumulate by adding the ordinates, one at a time between $h = 1.25$, and $h = 0$, moving backwards in h and plotting as we go. The process terminates in $N(0, 850) = 7/3 = 2.333$. This procedure is repeated for all the curves resulting in the set of points (○) shown in Fig. A.6.

The curves in Fig. A.7 are computed from (A.9), and the solid dots are taken from the line $h_0 = 0.4$ in Fig. A.6. The reasons for showing these alternative ways of arriving at the same distribution will become obvious in A.10 when dealing with

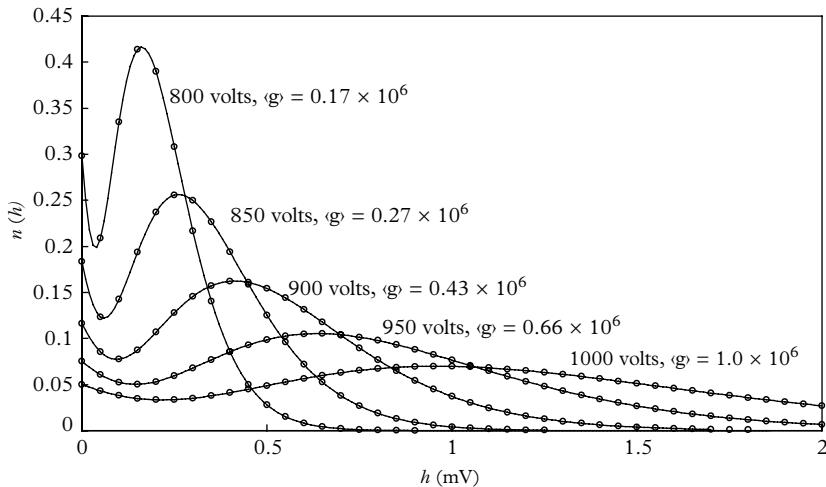


Fig. A.5. Differential distributions, shown as continuous curves, are derived from (A.8) for the operating voltages indicated. Because the light level is constant, the area under all curves must be common. Gain, $\langle g \rangle$, is also shown against each curve. The data points (\circ) are used to derive the integral distributions of Fig. A.6.

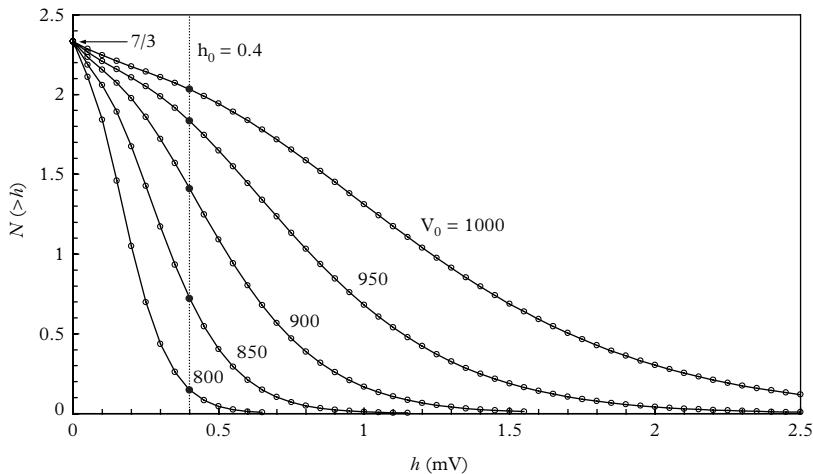


Fig. A.6. Integral pulse height distributions computed from (A.9) are shown by the solid lines. The circles refer to the numerical integration derived from Fig. A.5. The significance of the intersection of the vertical dashed line with the family of curves is discussed in the text.

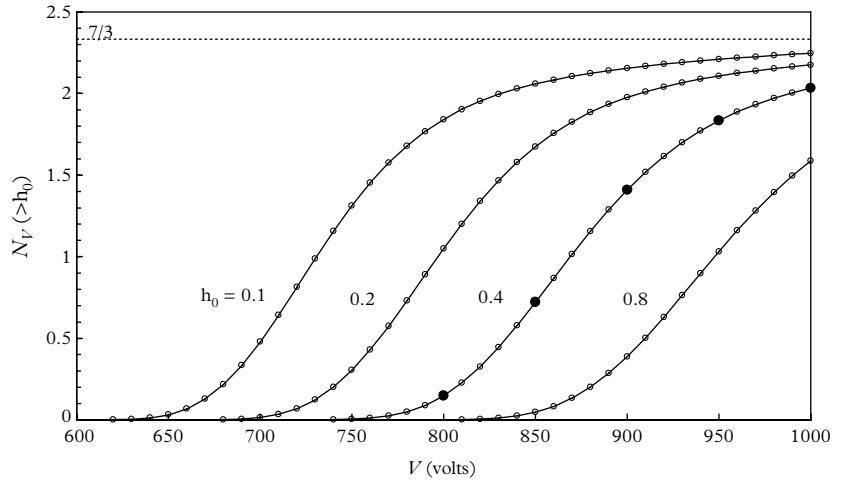


Fig. A.7. Plateau characteristics corresponding to the discriminator levels, h_0 , are indicated. The curves are the computation of (A.9). The five marked data points (●) on the line $h_0 = 0.4$ are taken from the vertical dashed line of Fig. A.6 as a check.

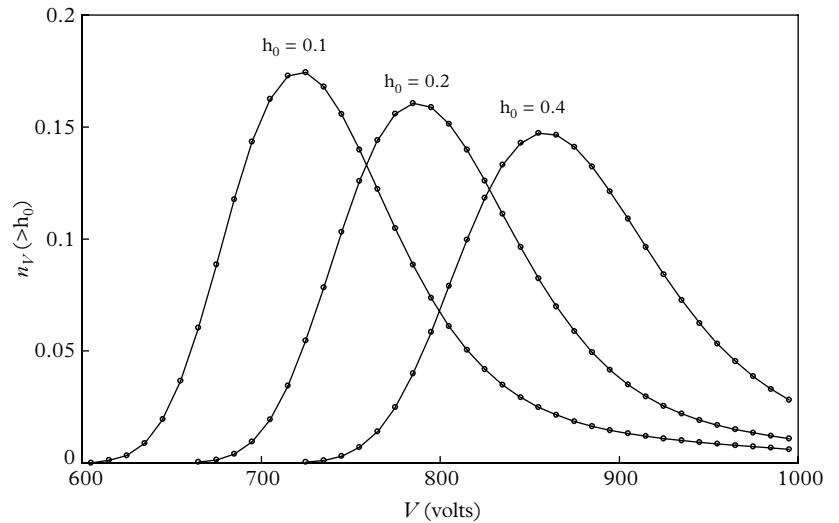


Fig. A.8. Differential pulse height distributions as a function of V . These curves provide a useful guide for selecting the operating point on a plateau characteristic, particularly where photon counting is concerned. The area under all distributions is common.

experimental data. The curves in Fig. A.7 are referred to as ‘plateau characteristics’, acknowledging the observation that, at sufficiently high gain or, equivalently, HV, such curves appear to flatten off and tend to the same asymptotic value (in this example, 7/3). We may say, from a practical point of view, that, at sufficiently high gain, most output pulses already exceed the threshold, and any further increase in HV will result in little increase in counts. This idyllic situation is not attained with all sources or with all PMTs. However, the usage of the word ‘plateau’ persists.

The differential density distribution, $n(h_0, V)$, expressed as a function of V , completes the set of curves (Fig. A.8). They may also be derived from tabular readings of $N(>h_0, V)$ by subtracting consecutive entries of $N(>h_0, V)$, from which $\Delta N(>h_0, V)/\Delta V \approx n(h_0, V)$ follows. This method is particularly useful for experimental results, provided that they are sufficiently closely spaced.

It may appear that all the curves in Fig. A.7 are parallel; however, this is only approximate, as may be verified by shifting them along the abscissa to overlay one another. This procedure provides a convenient diagnostic tool in photon counting, for example, given a threshold detector and a PMT, both of variable sensitivity. Any gross deviations in the shape of the plateau characteristics could indicate contributions from afterpulses, double pulsing, or the onset of breakdown.

A.8 Transformation of experimental data

Theory has been developed for continuous variables but PMT spectral data are always discrete, prompting the question of how to deal with such information. The answer is, exactly as for continuous data. We start with a pulse height distribution, measured at $V_0 = 1000$ V and listed in the first two columns of Table A.1. The data consist of ten rows, to illustrate the methodology of generating the other three distributions. The sum of the readings in the second column is the integral of $n(h, V_0)$. In this example, $N(>h_0, V_0) = 7502$ counts per unit time for $h_0 > 20$; this is also the first reading in the third column. The remaining entries are obtained by subtracting the adjacent $n(h, V_0)$ values progressing to the end of the column. The log slope is taken as $\beta = 8$ and representative of a typical 10- to 12-stage PMT. The initial reading in the fourth column is 1000 V, and the next is $V = 1000 (20/40) = 917$ V, and so on. A differential pulse height distribution, $n(h, V_0)$ vs h , is plotted in Fig. A.9, together with its integral counterpart, $N(>h, V_0)$. The ordinate values for $N(>h_0, V)$ are those given in Column 3, and they are plotted at the corresponding V values in Column 4. The $n(h_0, V)$ entries in the last column are calculated from the approximate differential $\Delta N(>h, V_0)/\Delta V$ times the channel width Δh . The values for the abscissa, V' , are taken from Column 5 and lie between pairs of V values. These readings are recorded in Fig. A.10.

Table A.1. Illustrating the tabular method for deriving the set of three distributions from the differential pulse height distribution listed in the first two columns. The voltages V' are located midway between V readings and they are required for plotting the differential $n(h_0, V')$. The small discrepancy in the two sums stems from the relatively wide increment taken for Δh .

h	n(h, V_0)	N(>h, V_0)	V	V'	n(h_0, V')
20	742	7502	1000		
40	510	6760	917	959	179
60	680	6250	872	894	225
80	1300	5570	841	856	442
100	1640	4270	818	829	1124
120	1280	2630	799	809	1780
140	740	1350	784	792	1678
160	370	610	771	778	1140
180	170	240	760	766	657
200	<u>70</u>	70	750	755	<u>342</u>
Sum				Sum	
7502				7567	

A.9 Single-photoelectron distributions

A plateau characteristic is the preferred basis of operation for commercial applications of PMTs. The foregoing theory is applied to single-photoelectron distributions, to be followed by investigations of multi-photoelectron pulse height distributions. The pulse height distribution shown in Fig. A.11 is typical of the SER curves for PMTs (also see Fig. A.15). There are two ways in which to express the quality of a PMT for such applications. First, it may be expressed in terms of the variance of the distribution; second, it may be expressed as the ratio of the peak height to the valley (P/V) of the SER. The variance involves calculation, whereas determining the P/V ratio is easily done from the ratio of just two ordinate values. It will become obvious why the P/V ratio serves as a better indicator of the

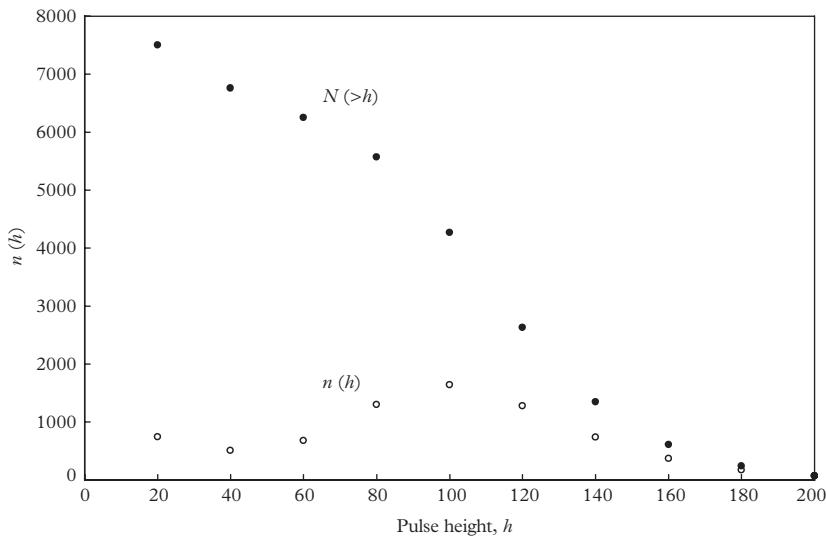


Fig. A.9. Data from Table A.1. The integral distribution is derived from the measured differential distribution starting at $h = 200$ and stepping to the left.

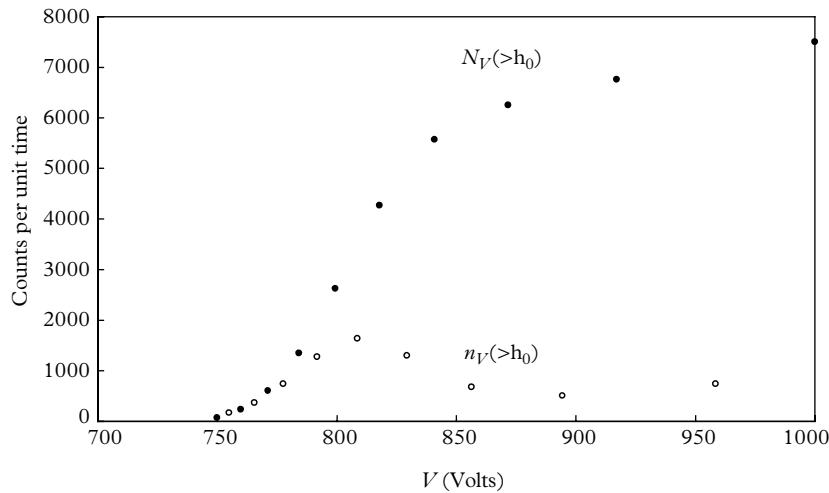


Fig. A.10. Transformation of discrete data, based on Table A.1.

photon-counting quality of a PMT, rather than the variance. The P/V ratio of this PMT is less than 2, whereas with selection it is possible to obtain samples where the P/V exceeds 3. The gain–voltage curve for this PMT, shown in Fig. A.12, provides the relationship between V and h ; we seek to make the transition from $n(h, V_0)$ to $N(>h_0, V)$ —the plateau characteristic.

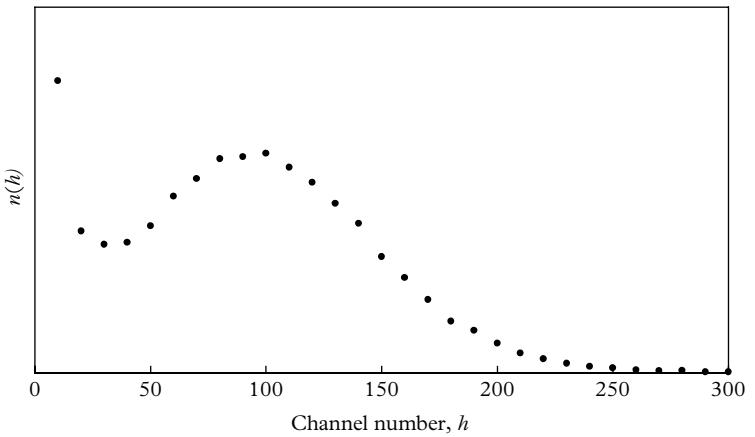


Fig. A.11. SER curve, for a 30 mm PMT measured at $V_0 = 1100$ V.

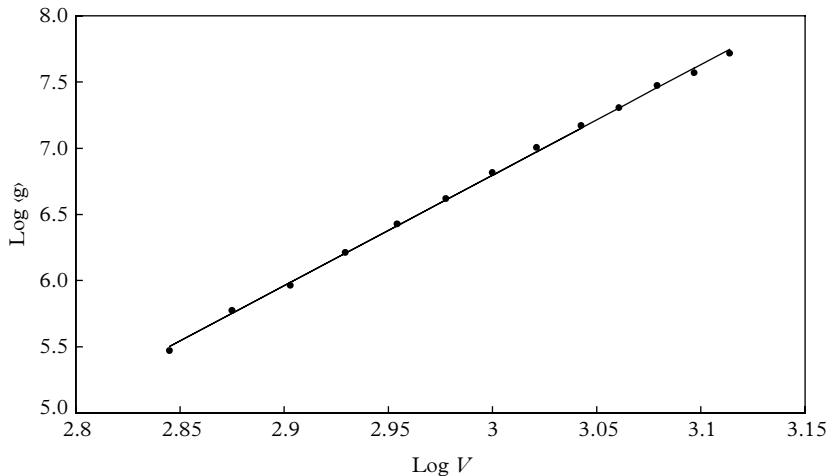


Fig. A.12. Measured gain–voltage curves for the PMT selected for Fig. A.11, using the same designation for the data points. The straight line is $h = 0.604 \times 10^{-18} V^{8.34}$. Note that h is synonymous with gain.

Methods for determining gain are described in 5.8 but the important parameter insofar as transformation is concerned is the exponent β , which is 8.34 for the PMT of Fig. A.12. A plateau characteristic measured with a fixed threshold discriminator is represented by the solid dots in Fig. A.13. Taking the plateau counts as $\sim 20,000$, indicated by the upper arrow, half this rate is found at 840 volts—shown by the vertical dashed line. The half-plateau count rate

corresponds to the median of the SER and we may, for present purposes, take this as 1 photoelectron equivalent. An approximate scale is easily derived using the gain–voltage relationship in Fig. A.12. This is attached to the upper x-axis, highlighting the contribution made by small pulses as V moves to the right. The differential distribution suggests an operating point of $V = 1000$ V, where $n(h_0, V)$ starts to flatten out at 1000 counts.

Photon-counting plateau characteristics are invariably plotted using a logarithmic scale on the ordinate, for the comfort that a flatter-looking curve provides. The validity of this statement becomes obvious when comparing Fig. A.13 and Fig. A.14, but note that the flatter appearance of the plateau is an illusion created by the use of a log scale. There is a tendency for users of PMTs to operate at a higher gain than is justified. The rationale for higher gain is twofold: first, statistical considerations always favour high counts; second, setting the operating point far to the right of the knee guards against the effects of gain loss. Why this may not be a good strategy can be seen by superimposing a photoelectron scale on the plateau characteristic. This scale is derived from (A.5) and we see that, when operating the PMT in the vicinity of 1500 V, pulses as small as 1/100th of a photoelectron equivalent contribute to the total: this should be of concern, warranting some justification.

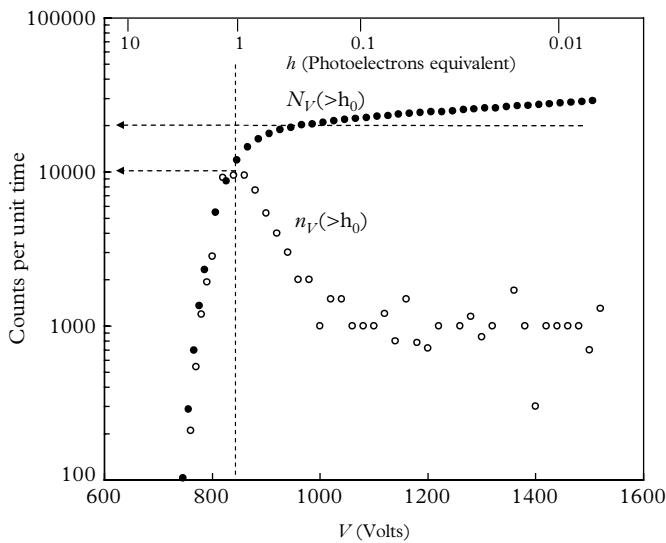


Fig. A.13. A plateau pulse height distribution $N_V(>h_0)$ (●), from which the differential, $n_V(>h_0)$ (○), derives by subtraction of adjacent plateau points. The slope of the plateau curve at any point is given by $n_V(>h_0)$. Note that $n_V(>h_0)$ attains a maximum at the position of the single-electron peak—as it must do.

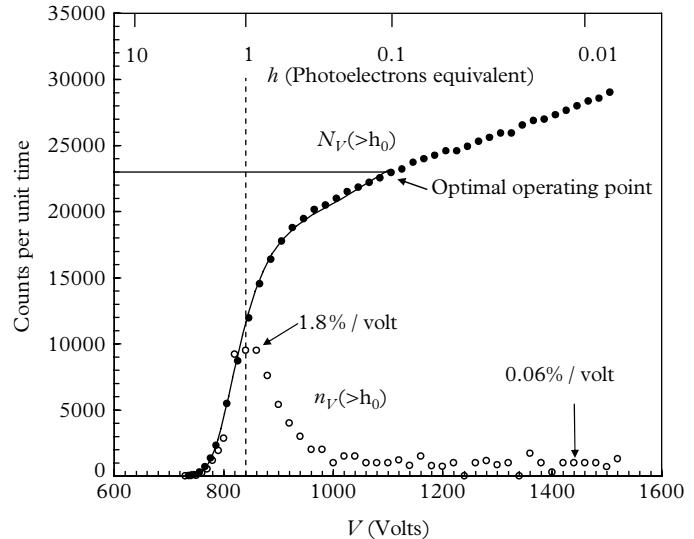


Fig. A.14. The data in these distributions is the same as that in Fig. A.13—only the ordinate scale is different. The measured plateau characteristic, $N_V(>h_0)$ (●), is for the PMT with the SER given in Fig. A.11. The curved line is a transformation of the differential spectrum taken with an MCA but note that the curve stops at 1150 V because, beyond this voltage, events increasingly fall outside the pulse height range of the MCA, and are unaccounted. Operating at 1000 V gives a count-rate stability of 0.06 %/volts, as indicated. The open circles suggest operating in the region of 1100 V.

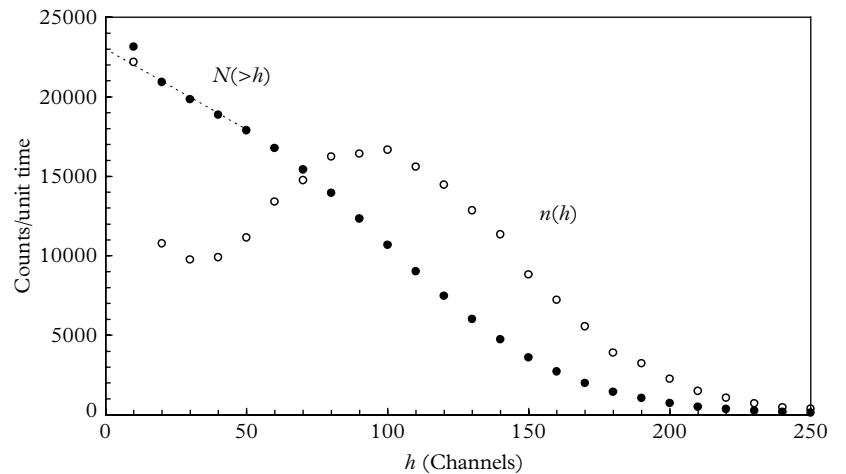


Fig. A.15. Integral and differential distributions taken from Fig. A.11 and measured at $V_0 = 1100$ V. The differential distribution has been scaled by 10. Extrapolation to $h = 0$ is shown by the dotted line, terminating at 23,000 counts.

A.10 Probability distributions in high temperature NaI(Tl) applications

Pulse height distributions for a NaI(Tl) crystal excited by ^{137}Cs gamma rays are shown in Fig. A.16. There are two spectra: one for operation at 20°C, and the other at 160 °C, both taken at the same HV. Loss in amplitude with temperature is caused by a reduction in the light output from the crystal, a loss of photocathode sensitivity, and a decrease in multiplier gain. Resolution consequently deteriorates at the higher temperature. The ordinate is the variable $n(h)|_{V_0}$, and the abscissa is h , expressed in keV. The gain–voltage relationship is derived using the variation in the NaI(Tl) peak position with HV, at the two temperatures in Fig. A.17.

The plateau characteristic shown in Fig. A.18 is generated from the differential spectrum shown in Fig. A.16, $n(h, V_0)$, following the procedure described in A.8. The ordinates are obtained from the tabulated spectral values, and are given by

$$N(>h, V_0) = \sum_{h=h_0}^{1020} n(h, V_0), \quad (\text{A.10})$$

where V_0 is the HV at which the pulse height distribution was measured. The lowest usable channel of the MCA is $h_0=12$; the instrument has a built-in discriminator and any channel data below 12 is distorted. The data presentation therefore starts at $h=12$ and extends to Channel 1020, the upper limit to the range of the analyser. The plateau voltages corresponding to each h value are given by $V = V_0 (h_0/h)^{1/\beta}$, from (A.5). The integral distribution, $N(>h_0, V)$, is derived from

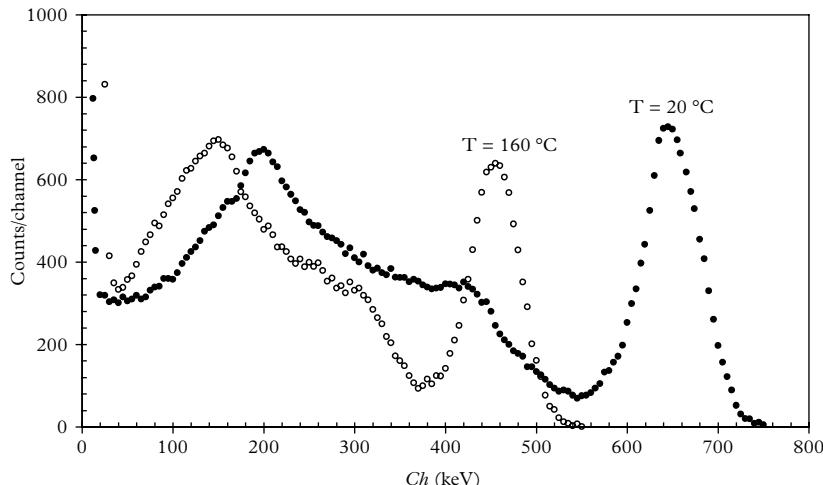


Fig. A.16. Differential pulse height distributions for NaI(Tl) at the temperatures indicated. MCA sensitivity is 1 channel/keV. The PMT is a 12-stage 9224 metal–ceramic type.

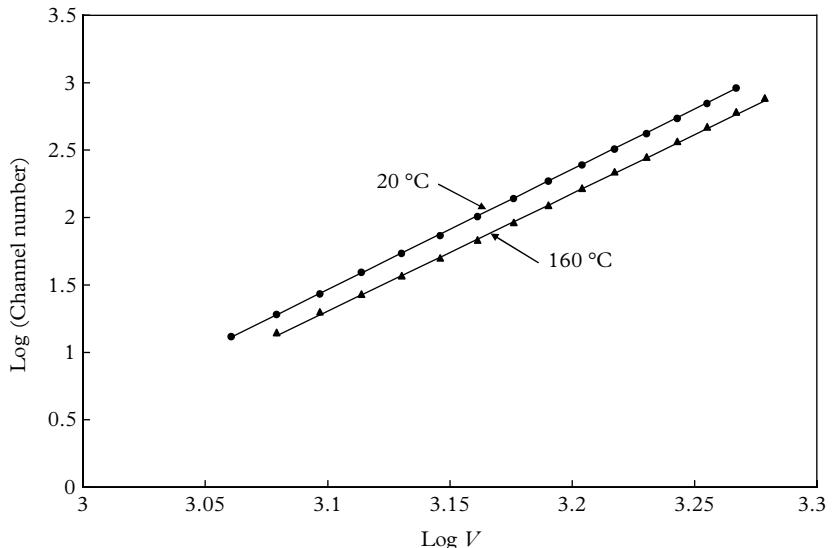


Fig. A.17. Gain–voltage relationships for the PMT in Fig. A.16, of the form $h = aV^\beta$. The ordinate is the channel number of the MCA at which the ^{137}Cs peak occurs. Parameters take the following values at 20 °C, $\beta = 8.93$; at 160 °C, $\beta = 8.72$.

$N(>h, V_0)$ data using the same ordinate values but replacing abscissa values, h , with corresponding V values in accordance with $V = V_0(h_0/h)^{1/\beta}$, as shown in Fig. A.18 by the thin line. The small dots represent the plateau characteristic measured with the MCA readings treated as an integral discriminator, with $h_0 = 12$. This is done using the region of interest facility, set to integrate all channels between Channel 12 and Channel 1020. Data can be accumulated in this way until the magnitude of the HV is such that some signals exceed the maximum channel and are lost. The third set of data is indicated by circles, with a threshold discriminator comprising the same type of preamplifier and main amplifier used with the MCA. In this case, the main amplifier is connected to a fast 50 mV threshold NIM discriminator. The sensitivity of the threshold discriminator is set to that of the MCA using the method described with reference to Fig. A.13. It is gratifying, but not surprising, that the plateau characteristic measured in this way aligns with the other two, and extends to higher voltages without loss of counts. The mini plateau at around 1150 V in Fig. A.18 corresponds to the ^{137}Cs peak, which can serve to calibrate the abscissa in terms of keV. This feature is usually lost when recording a plateau characteristic by using a standalone discriminator and 20 V steps. The derivation of the plateau curve at 160 °C (Fig. A.19) is measured in the same way, while the integral plots in Fig. A.20 follow from integrating Fig. A.16.

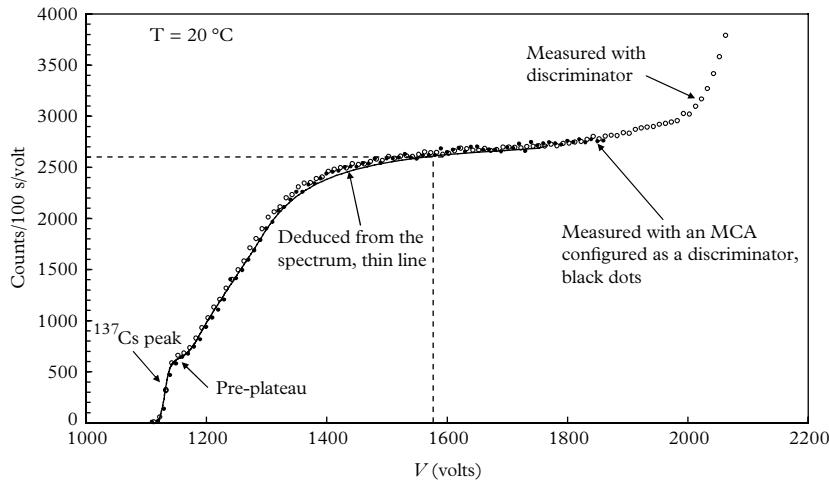


Fig. A.18. Plateau characteristic at 20°C , measured by three different means. The identification of the ^{137}Cs peak and its associated pre-plateau is marked in the figure. Integration of the photopeak in the differential distribution gives rise to the pre-plateau and the peak position, as indicated, serving as a means of energy calibration.

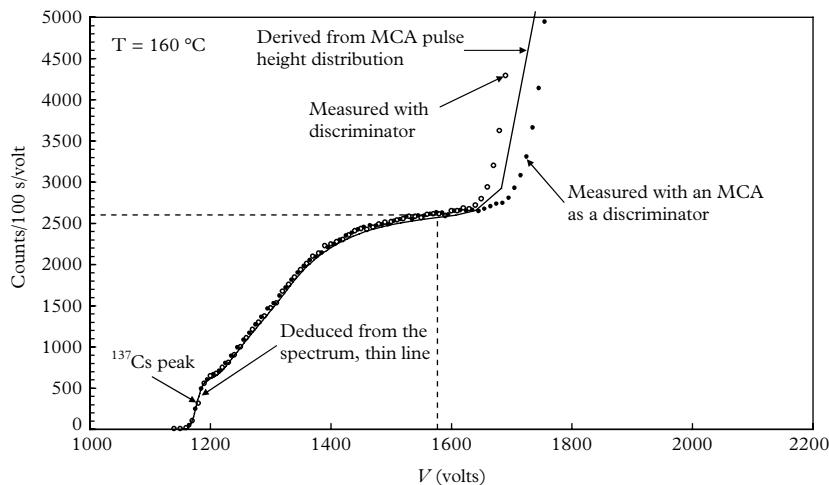


Fig. A.19. The plateau characteristic at 160°C , measured by three different means.

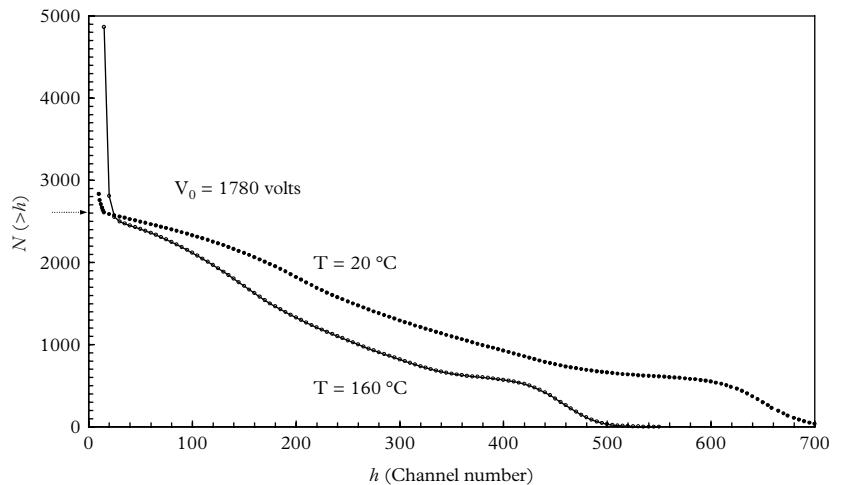


Fig. A.20. Integral distributions corresponding to the spectral distributions in Fig. A.16. Extrapolation to $h = 0$ indicates a total count rate of 2600 for both sets of readings.

A.11 Conclusions

It has been shown that, given one of the four PMT distributions and the corresponding gain–voltage characteristic, it is possible to derive the other three. An MCA is the most efficient means of gathering data, since every event within its range is analysed. A differential spectrum can be measured with an SCA, but it has to be created one channel at a time, and this takes up to 1024 times longer than required with an MCA. Data collection takes as long by this method, as with an SCA.

Apart from the inefficiency of data collection, there are also statistical limitations on the precision of the data. In applications with an MCA or a SCA, and in which Poisson statistics are assumed to apply, the one-sigma uncertainty in the counts n in any channel is \sqrt{n} . A differential distribution, derived by subtracting adjacent readings taken with a fixed threshold discriminator, will always be noisier than the one measured directly with an SCA or an MCA, especially in the regions where the spectrum is flat, in which case the standard deviation on individual differential readings will vary as $\sqrt{(2n)}$, and the noise may exceed the reading itself.

Plateau measurements are typically made in steps of 20 to 50 V, and much of the detail shown Figs A.18 and A.19, for example, is suppressed. The practical determination of a plateau curve, based on measuring counts $>h_0$ over a range of voltages may highlight inconsistencies which show up when a repeated measurement is made at some earlier voltage setting—the two measurements seldom agree within the expected statistical uncertainty. This anomaly is also present, for example, if more detail is sought between an existing pair of readings: the finer detail thus measured may be corrupted by short-term instability because the PMT

gain has not been allowed sufficient time to stabilize at each measurement point. The inconsistencies that arise in a repeated measurement, made, say, 15 min later than the first one, are usually attributed to hysteresis effects.

The recommended measurement technique is one of recording a preliminary plateau characteristic, in relatively coarse steps, to establish the appropriate range for V ; a compatible voltage increment, ΔV ; and an appropriate time interval for the count-rate determination. Where signal measurement is concerned, it is often possible to avoid the statistical limitations referred to previously by counting for only 1 to 10 s per data point; background measurements, however, may need longer counting intervals if their rates are low. The final measurement of the plateau characteristic should be done in an ordered way, avoiding interruptions or backtracking. Ideal conditions usually apply in commercial instrumentation in which the plateau characteristic is measured automatically but those done by hand are all too often subject to interruption with the aforementioned consequences.

It is obvious that we have been dealing with well-behaved distributions, and with ideal PMT gain behaviour. In reality, pulse height distributions, such as those shown in Fig. A.11, change with increasing HV because of signal-induced noise; afterpulses and dark counts tend to increase. Gain non-linearity is also a factor, and β may decrease in some PMTs at high gain. The decision faced in any application involving a plateau characteristic is the one of where to set the operating point. Precise setting is often a question of personal preference, but the presentations in Figs A.19 and A.20 offer guidance, provided the user is clear on precisely what is to be measured—which is not always the case. There is a contribution to the signal plateau characteristic from background—in fact, the background has its own plateau characteristic. The shape of the background plateau distribution is always different from that due to the signal; it tends to be more structured because of the multiplicity of sources involved. Plotting the ratio of the signal to background plateau counts, as a function of HV, sometimes helps in selecting the optimum operating point. The magnitude of the background contribution is defined once this point has been chosen and it may be subtracted from any subsequent signal plus background measurement.

An important requirement is that the slope of the plateau count rate should be positive everywhere on the curve, excluding the effects of statistical fluctuations. Failure to satisfy this criterion suggests excessive dead time, voltage divider instability, or some other malfunction. A plateau count rate $N(>h_0, V)$ approaches the integral count rate, $N(>h, V_0)$ for $V \rightarrow \infty$, corresponding to $h = 0$. This is evident in Fig. A.6 and Fig. A.7 and it raises an obvious practical problem in choosing the appropriate operating point on a sloping plateau curve. Referring to Fig. A.19, one may wish to relate the selected plateau count rate to an integral rate of Fig. A.20. Note that experimental distributions cannot register a reading at $h = 0$. Perhaps the best that can be achieved, in the presence of a pedestal, is to extrapolate the integral curve to zero pulse height, as is evident in Figs A.15 and A.20. This particular uncertainty affects the determination of collection efficiency, and commercial instruments based on photon counting.

Appendix B

Light emission by the Cerenkov effect

B.1 Introduction

The Cerenkov effect is a source of PMT background of importance in many applications—arguably, the phenomenon was first measured by Cerenkov by using a PMT. The purpose of the sections that follow is to provide the necessary formulae for the computation of light yield in this process. The well-known angular distribution of the light and its strong dependency on wavelength are also covered. As with many theories, there are always refinements, but it is sufficient for our purposes to investigate only the basic formulation. The first consideration is to present formulae for the energy required, of electrons, to initiate Cerenkov emission; this is pertinent to relatively low-energy electrons produced by gammas with energy in the MeV range. No attempt is made to calculate the total yield for any given low-energy electron in coming to rest. It is sufficient for our purposes to know that gammas from ^{57}Co , in glass, for example, cannot produce Cerenkov emission, whereas those of ^{137}Cs can. Gamma rays give rise to Cerenkov light, but only indirectly. A relativistic electron must first be produced by any one of the electromagnetic interactions, and it is this electron that produces light. The light yield from MIPs that remain relativistic and traverse the full window thickness of a PMT is calculated to compare with actual measurement.

Energetic, charged particles such as electrons, protons, and muons produce light when passing through transparent material. The light itself is derived from various energy loss processes: there is a continuous loss due to excitation and ionization of the medium surrounding the track of a particle, together with the occasional transfer of a significant fraction of the energy, through such processes as knock-on, bremsstrahlung, and pair production. Visible radiation generated by ionization is usually referred to as fluorescent, phosphorescent, or, scintillation light, but only one photon is released per interaction. Relativistic charged particles, with energy beyond a prescribed threshold, emit light directly through the Cerenkov effect. This process is examined in some detail, to relate the energy and the refractive index of a medium to the quantity of light produced.

B.2 Threshold conditions for the Compton effect

In Compton scattering, an incident photon of energy $h\nu$ imparts some of its energy to an atomic electron. The recoil electron acquires kinetic energy in this process, subject to the rules of conservation of energy and momentum. A scattered photon, $h\nu'$, is always produced with energy $h\nu' < h\nu$, the incident energy. Compton scattering is well documented in the scientific literature, and the treatments by Evans (1955) and Knoll (1979) are particularly useful.

The energy of the scattered photon is given by (B.1), and the electron energy is given by (B.2):

$$h\nu' = \frac{h\nu}{1 + \alpha(1 - \cos \theta)} \quad (\text{B.1})$$

$$T = h\nu \frac{\alpha(1 - \cos \theta)}{1 + \alpha(1 - \cos \theta)}, \quad (\text{B.2})$$

where θ is the angle between the path of the incident photon and that of the scattered photon, and α is defined as $h\nu/m_0c^2$, where m_0 is the rest mass of the electron. According to (B.1), $h\nu'$ attains its minimum value when $\theta=180^\circ$ (back-scattering). Consequently, the Compton electron carries away the maximum energy that is kinematically possible when $\theta=180^\circ$. Equations (B.1) and (B.2) are plotted in Fig. B.1 for $\theta=180^\circ$. The energy of the backscattered photon approaches $m_0c^2/2=0.255$ MeV, for incident photons of sufficiently high energy.

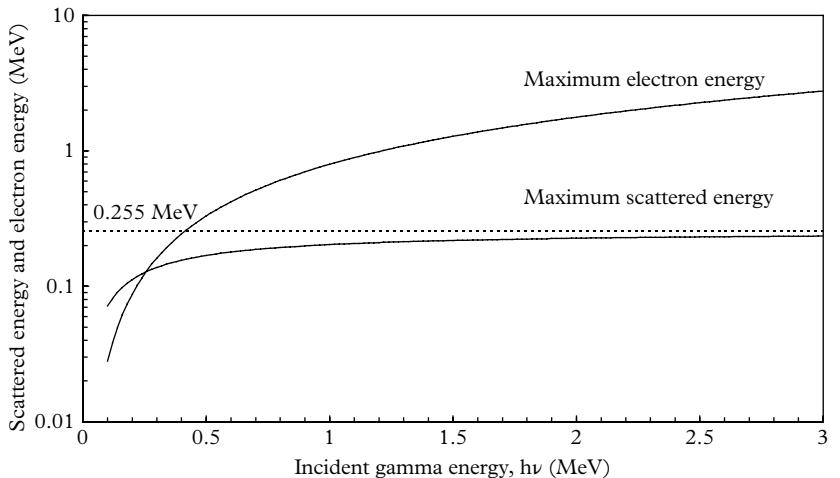


Fig. B.1. The maximum electron energy produced by Compton scattering is obtained for $\theta=180^\circ$. The electron energy approaches that of the incident photon, for all $h\nu \geq 1$ MeV. The maximum energy of the scattered photon tends to $m_0c^2/2=0.255$ MeV for gammas with sufficiently high incident energy, as indicated by the dashed line.

In the Compton effect, any energy between 0 and $h(\nu - \nu')$ may be transferred to an electron, with the maximum subject to (B.2). The requirement for the onset of light emission by a charged particle with speed ν is $\beta n > 1$, where $\beta = \nu/c$, and n is the refractive index of the medium. Starting with Einstein's relationship between the total energy, E , and β , we have the following condition for the onset of light emission at $\beta n = 1$:

$$E = \frac{m_0 c^2}{(1 - \beta^2)^{1/2}} = \frac{m_0 c^2}{(1 - n^{-2})^{1/2}} = m_0 c^2 \left(\frac{n^2}{n^2 - 1} \right)^{1/2} \quad (\text{B.3})$$

We then find that the energy imparted to the electron is

$$T = E - m_0 c^2 = m_0 c^2 \left[-1 + \left(\frac{n^2}{n^2 - 1} \right)^{1/2} \right], \quad (\text{B.4})$$

remembering that T refers to the electron energy satisfying $\beta n > 1$, and only those electrons with energy in excess of this threshold can emit photons. Equation (B.4) and the curves shown in Fig. B.2 are the same as those given by Sowerby (1971).

The photon energy necessary to produce an electron of energy T follows from (B.4), and with $\theta = 180^\circ$, as before,

$$h\nu = T/2 + \frac{1}{2}(T^2 + 2m_0 c^2 T)^{1/2} \quad (\text{B.5})$$

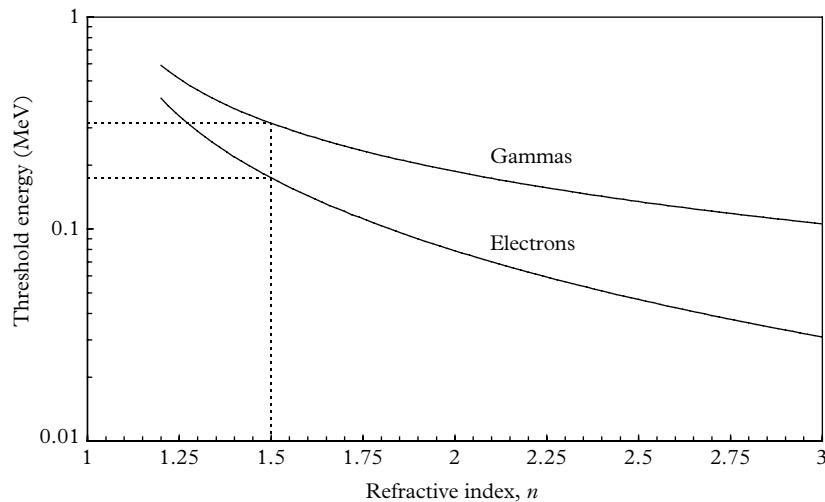


Fig. B.2. The threshold energy for Cerenkov emission as a function of n , the refractive index of the medium. The upper curve is the minimum energy gamma ray (0.316 MeV for $n=1.5$) just capable of imparting sufficient kinetic energy, T , through Compton scattering, to generate light. The lower curve represents the minimum energy of an electron for the onset of Cerenkov radiation (0.175 MeV for $n=1.5$).

This relationship is plotted in Fig. B.2 in the upper curve. The index of refraction for typical window materials is 1.5, for which the electron threshold energy is 0.175 MeV; the corresponding minimum gamma energy is 0.316 MeV, as indicated in B.2. In other words, an electron of energy >0.175 MeV will emit Cerenkov photons in a medium with refractive index 1.5—for example, glass. Emission ceases once the electron has decelerated below this threshold. An electron with energy of 0.175 MeV can only be produced by a gamma ray of energy >0.316 MeV. The naturally occurring gamma background extends up to 2.615 MeV—an isotope of ^{208}Tl . Consequently, a proportion of this gamma continuum meets the energy requirements for the emission of light. All gammas of energy greater than 0.316 MeV, whether from naturally occurring radionuclides or otherwise, will contribute to PMT background through the combination of the Compton and Cerenkov effects. All gammas of energy in excess of 0.175 MeV will generate light via the combination of the photoelectric and Cerenkov effect, or through pair production and the Cerenkov effect.

B.3 Formulations for Cerenkov emission

Frank and Tamm (1937) showed that the energy, dE , radiated in a short distance dx by an energetic charged particle in traversing a transparent medium is given by the expression in (B.6). Theirs was the classical derivation of the Cerenkov formula, but the quantum mechanical treatment gives the same result:

$$\frac{dE}{dx} = \frac{2\pi z^2 \alpha h}{c} \left[1 - \frac{1}{(\beta n(\lambda))^2} \right] \int_{\nu_1}^{\nu_2} \nu d\nu \quad \text{J} \quad (\text{B.6})$$

The integration is taken over all frequencies, $\nu = c/\lambda$, for which the refractive index $n(\lambda)$ exceeds unity. The refractive index approaches unity for those wavelengths that are strongly absorbed, and photon production ceases. Cut-off occurs in the UV for the window materials commonly used in PMTs. For relativistic particles of unity charge, we have both $\beta=1$, and $z=1$. The fine structure constant is $\alpha=1/137$, h is the Planck constant, and c is the speed of light in vacuum. The opening angle of the light in glass is given by $\cos \theta = 1/\beta n(\lambda) = 48^\circ$, which is not very different from the critical angle of 42° . The significance of this observation is discussed in 6.8.

Equation (B.6) gives the energy radiated in joules within the frequency band ν_1 to ν_2 . The refractive index is essentially constant over the visible band but it increases significantly for UV light, as shown in Fig. B.3.

We can express (B.6) in terms of wavelength by performing the integration, which leads to

$$\frac{dE}{dx} = \frac{\pi \alpha h c}{e} \sin^2 \theta \left(\frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2} \right) \text{ eV/m.} \quad (\text{B.7})$$

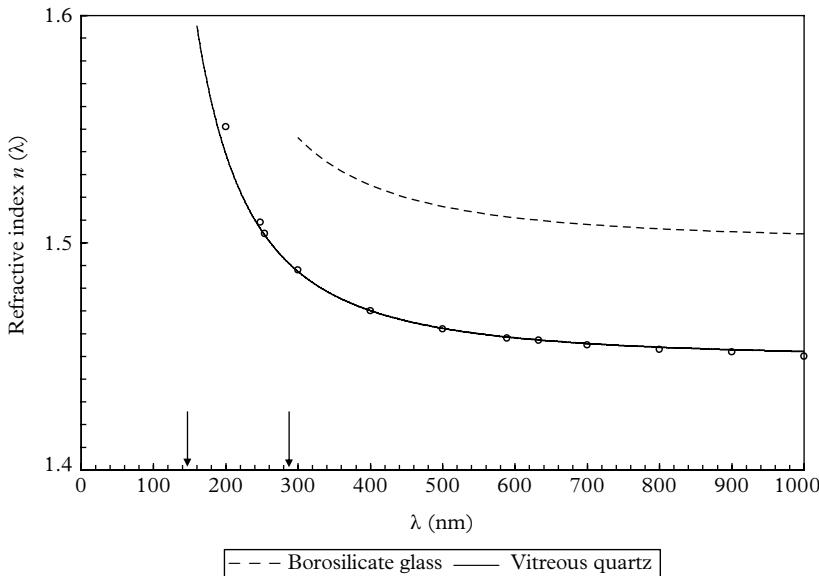


Fig. B.3. Variation of refractive index with wavelength for window materials used in PMTs. The curves are derived by application of the Cauchy formula given in Jenkins and White (1957) and using data from Malitson (1965). The arrows indicate the cut-off wavelengths of the two materials.

The units take the form of eV per metre since we have divided by the electronic charge, e . We return to (B.6) to derive an expression for the number of quanta emitted per unit wavelength, which is

$$\frac{dE}{dx} = \frac{\pi\alpha h}{c} \sin^2\theta (\nu_1^2 - \nu_2^2) \quad (B.8)$$

Consider the average energy radiated between two frequencies ν_1 and ν_2 , which is $\bar{\nu} = (\nu_1 + \nu_2)/2$. We seek an expression, in terms of wavelength, for the average number of quanta emitted per metre:

$$\begin{aligned} \frac{1}{h\bar{\nu}} \frac{dE}{dx} &= \frac{1}{(\nu_1 + \nu_2)/2} \frac{\pi\alpha h}{c} \sin^2\theta (\nu_2^2 - \nu_1^2), \\ \frac{1}{E} \frac{dE}{dx} &= \frac{2\pi\alpha}{c} \sin^2\theta (\nu_2 - \nu_1) \\ &= 2\pi\alpha \sin^2\theta \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right). \end{aligned} \quad (B.9)$$

Evaluating (B.7) and (B.9) for glass, we have, for the visible spectrum 400 to 800 nm, $dE/dx = 750$ eV/cm (compare this with the energy loss of 3.8 MeV/cm for

ionization and excitation losses) and $(1/E) dE/dx = 320$ photons/cm. The yields for other materials are easily calculated by choosing the appropriate value for $\sin^2\theta$. For a PMT with a mean QE of, say, 20 %, we predict about 6.4 pe per millimetre path length.

B.4 The yield from MIPs

In order to determine the quantity of photoelectrons more accurately, we need (B.9) in the form of the number of quanta per unit wavelength interval, $d\lambda$. We also require the photocathode QE (Fig. B.4) and the refractive index as functions of wavelength (Fig. B.3).

Let $\lambda_1 \rightarrow \lambda_2 = \lambda$; then, from (B.9),

$$\frac{1}{E} \frac{dE}{dx} = \frac{2\pi}{137} \sin^2\theta \frac{d\lambda}{\lambda^2} \quad \text{photons/m.} \quad (\text{B.10})$$

This expresses the well-known result that the number of photons emitted per unit wavelength is proportional to $1/\lambda^2$, and the light therefore appears bluish-white. This is the differential light emission expression for the number of quanta emitted between λ and $\lambda + d\lambda$. All transparent media are characterized by a particular UV cut-off, beyond which transmission ceases. The increased yield at short wavelengths, embodied in the term $1 - (n(\lambda))^{-2}$ in (B.6), is shown in Fig. B.3. The refractive index approaches unity once self-absorption sets in and Cerenkov emission ceases. Allowing for the change in $n(\lambda)$ with λ makes a few per cent

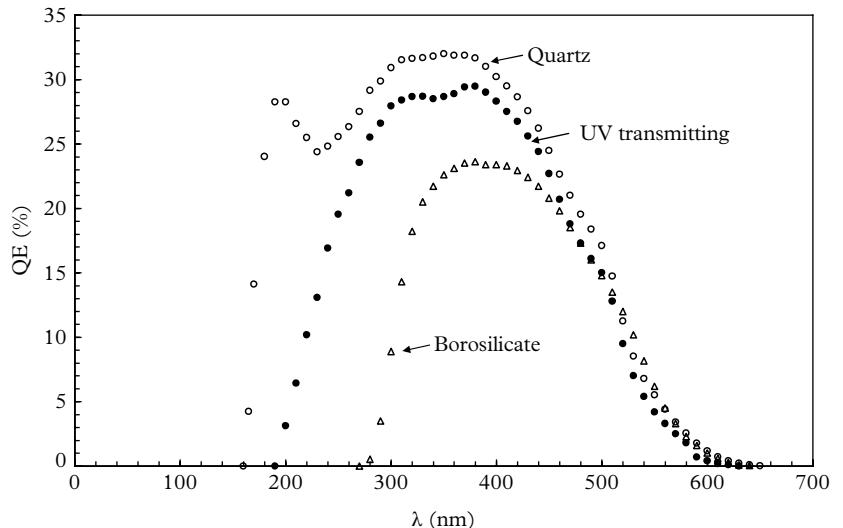


Fig. B.4. Quantum efficiency for the three window materials used with bialkali photocathodes.

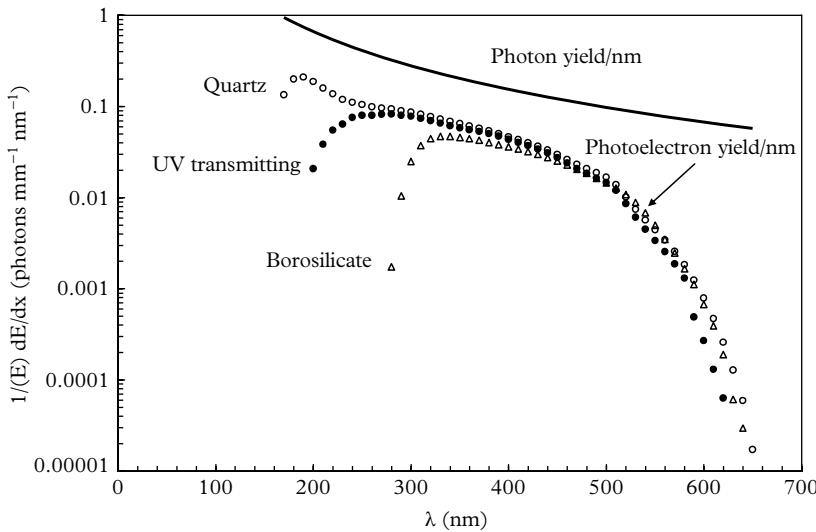


Fig. B.5. The upper curve is the photon yield, expressed in photons per millimetre per unit wavelength (nm). The lower curves, scaled by $\eta(\lambda)$, the QE, represent the photoelectron yield for bialkali photocathodes and the window materials featured.

difference to the realizable photon yield. Spectral sensitivity curves for PMTs with quartz, UV-transmitting glass, and standard borosilicate window materials are shown in Fig. B.4. The UV wavelength at which a photocathode first begins to show sensitivity is determined by the window transmission and not by the QE, and these cut-off wavelengths are indicated in Fig. B.3. We can determine the number of photoelectrons produced by folding the spectral sensitivity curve with (B.10), as shown in Fig. B.5.

The differential photon yield has been calculated from (B.10) for quartz and it is shown as the upper curve in Fig. B.5. Similar curves for the other materials are easily calculated but are not shown in the figure. Folding-in the QE curves for three spectral responses leads to the differential photoelectron yield, shown in the lower curves. The significant contribution to the photoelectron yield, derived from the UV part of the spectrum, is clearly evident in all three curves.

The information in Fig. B.5 predicts the size of the PMT signal in photoelectrons. This is given by the area under the lower curves of the figure. The photoelectron yields for the particular bialkali photocathodes, characterized by the QE curves of Fig. B.4, are as follows (in units of photoelectrons per millimetre of window thickness): quartz, 28; UV-transmitting glass, 16; and borosilicate, 7.5. These photoelectron yields have been verified experimentally in 6.8.1 by measuring the signals produced by cosmic ray muons in traversing a PMT window.

REFERENCES

- Evans, R. D. (1955). *The Atomic Nucleus*. McGraw-Hill, New York.
- Frank, I. and Tamm, I. (1937). Coherent visible radiation from fast electrons passing through matter. *CR Acad. Sci. USSR*, **14**, 109–14.
- Jenkins, F. A. and White, H. E. (1957). *Fundamentals of Optics*. McGraw-Hill, New York.
- Knoll, G. F. (1979). *Radiation Detection and Measurement*. John Wiley and Sons, Inc., New York.
- Malitson, I. H. (1965). Interspecimen comparison of the refractive index of fused silica. *J. Opt. Soc. Am.*, **55**, 1205–8.
- Sowerby, B. D. (1971). Cerenkov detectors for low-energy gamma rays. *Nucl. Instr. and Meth.*, **97**, 145–9.

Appendix C

Abbreviations

The following table lists the abbreviations used in this book

ACT	Air Cerenkov telescope
ADC	Analogue-to-digital converter
APD	Avalanche photodiode
ASDL	Acceleration spectral density level
ASIC	Application-specific integrated circuit
CAMAC	Computer-automated measurement and control
CCD	Charge-coupled device
CEM	Channel electron multiplier
CF	Constant fraction
CFA	Current-feedback operational amplifier
CFD	Constant fraction discriminator
CMS	Compact muon solenoid
CW	Cockcroft–Walton
DC	Direct current
DMM	Digital multimeter
DSP	Digital signal processing
ECL	Emitter-coupled logic
EMA	Extramural absorber
ENC	Equivalent noise charge
ENF	Excess noise factor
FADC	Flash analogue-to-digital converter
FET	Field-effect transistor
FPGA	Field-programmable gate array
fwhm	Full width at half maximum
G–M	Geiger–Müller
GEM	Gas electron multiplier

612 Abbreviations

HACR	High-accuracy cryogenic radiometer
HAPD	Hybrid avalanche photodiode
HPD	Hybrid photodetector
HPMT	Hybrid photomultiplier
HPK	Hamamatsu Photonics
HV	High voltage
JFET	Junction field-effect transistor
IEEE	Institute of Electrical and Electronic Engineers
LED	Light-emitting diode
LIDAR	Light detection and ranging
LVD	Low-voltage differential
LVDS	Low-voltage-differential signalling
MCA	Multichannel analyser
MCP	Microchannel plate
MCPPMT	Microchannel photomultiplier
MCS	Multichannel scaler
MIP	Minimum ionizing particle
MMIC	Microwave monolithic integrated circuit
MOSFET	Metal oxide semiconductor field-effect transistor
MPHD	Multi-photon pulse height distribution
MWD	Measurements while drilling
MRL	Maximum recommended level
ND	Neutral density
NIM	Nuclear instrument module
OP-AMP	Operational amplifier
PC	Photon counter
PDE	Photon detection efficiency
pdf	Probability distribution function
PIN diode	Semiconductor photodiode
PLP	Picosecond light pulser
PMT	Photomultiplier
POM	Polyoxymethylene
PSD	Pulse shape discrimination
PSDL	Power spectral density level
PTFE	Poly(tetrafluoroethene)

P/V	Peak-to-valley
p/z	Pole-zero
QDC	Charge analogue-to-digital converter
QE	Quantum efficiency, η
SCA	Single-channel analyser
SER	Single-electron response
SETR	Single-electron time response
SiPM	Silicon photomultiplier
SNO	Sudbury Neutrino Observatory
SPS	Spontaneous parametric scattering
SRS	Stanford Research Systems
TAC	Time-to-amplitude Converter
TAL	Technical attenuation length
TDC	Time-to-digital converter
TE	Transverse electric
TEL	Technical attenuation length
TIR	Total internal reflection
TM	Transverse magnetic
TPB	Tetraphenyl butadiene
TTL	Transistor-transistor logic
VFA	Voltage-feedback operational amplifier
VLPC	Visible light photon counter
VUV	Vacuum ultraviolet
WLS	Wavelength shifting
YAP	Yttrium aluminium perovskite

Index

Page numbers in italics refer to illustrations and tables.

A

anode signals 341–3, 346–50, 386
anode current 339–40
anode detection efficiency 415
arrival time distributions 175–80

B

background 258–78
see also dark counts/current, signal induced
cosmic rays 276–9, 277
dark counts/current 2, 21–2, 69, 122, 258–67
light leaks 464
leakage current 261–4, 308
light generation 265–6
radio nuclides 274–5

C

calibrated lamp 425–6
cascade processes 153–4
Cerenkov radiation 92, 101, 175, 266–7, 269, 272, 277, 425–6, 603, 606–9
charge output transducers 530
circuit analysis 544–7
current feedback OP AMPS 566–71
digital signal processing 529–30, 574–6
discrete component amplifiers 562–5
equivalent circuits 511–12, 538–41
multiple signal outputs 543–4
passive circuits 544–9
shaping amplifiers 559
voltage-feedback OP-AMPS 566–9
collection efficiency 411–29
bi-photonic field 424–5
Compton scattering 272–6, 604–6
current detection (DC) 284–8, 289–90, 302–3, 307–8

D

dead time 188–96
detectors 12–18

box-car integrators 293–4
lock-in detection 294–5
multichannel analysers(MCAs) 259–62, 303–4, 311–12, 561, 582, 600
ratemeter operation 304–5
response function 439–40
single and multiphotoelectron source 130
stop-start instrumentation 434–5
synchronous signal averaging 295, 296–7
dynodes and multipliers 203–8, 222, 225–37
box-and-grid 234, 235, 302, 337
channel electron multipliers (CEMs) 225, 237–9
circular focus 232, 234
coarse mesh
dynodes 229
dynode materials 387
dynode signals 507–9
edge effect 225
fine mesh 230
gating 515–24
inelastic scattering 217–21
linear focussed 232, 233
metal channel 230–2
Photonis foil 232
skipping 227–9
transmission 234–7
dynamic range 388–91

E

electric fields 458–61
electronics 528–75
amplifier discriminators 432–3
CAMAC 537
charge sensitive amplifiers 550–4
constant fraction timing (CF) 363–5, 376
digital multimeters (DMMs) 194, 246, 317–18
discriminators 365–6, 375–6
electrometers 318
ground connections 534

logarithmic amplifiers 571–2
microwave transistors 564–5
modules 535–8
nuclear instrument module (NIM)
units 537–8, 566
PMT-electronics interface 532–4
pole-zero cancellation 554
preamplifiers 531–4, 537–8, 549–59
pulse generators 522–3
shaping amplifiers 555–8
thyristors 517
timing-filter amplifiers 359–61
transimpedance amplifiers 569–71
electron motion in fields 322–34
electron-hole generation 204
environmental considerations 456–70
atmospheric helium 448–9
electric fields 458–61
exposure to bright light 449–52
high magnetic fields 458
high temperatures 68–9
ionizing radiation 72–3, 269–76
low temperatures 65–7
minimum ionizing particles (MIPS) 102, 608–9
shock and vibration 456–7
ultra low temperature 67–8
weak magnetic fields 456
excess noise factor (ENF) 7, 156, 164

F

fast timing 353–68
focus electrode gating 521
Fresnels equation 40, 42

G

gain determination 241–53
gain-voltage characteristics 252–3, 473–7
multiplier 245–9
photomultiplier 241–5
quick determination 251–2

- gamma rays 88, 105, 269–73, 274–6
 gating 515–24
 graded seals 275
- H**
 HV polarity and ripple 503–5
- J**
 jitter 331, 345–7, 356–61, 369
- K**
 Kerr effect 517
 Kubetsky, L A 3
- L**
 light guides 97–101, 109–14
 adiabatic 99–101
 attenuation length 80–1, 415–16
 diffuse reflectors 92–5, 102, 108
 escape cones 95–7
 fibres 100, 110–12
 fishtail 100
 hollow 98
 manifold 100, 101
 solid 98–9
 Winston cone 91–2
 wavelength shifting (WLS) 109–11,
 112–14
- M**
 microchannel plates (MCPs) 239–41
 double-chevron devices 241
- N**
 noise 2, 7–10, 161–5, 283, 559, 562
 detector noise 559–61
 determination of collection
 efficiency 422–4
 determination of
 non-linearity 403–4
 Johnson noise 5–6
 low frequency (1/f) noise 264
 shot noise power detection 286
 shot noise 161–3, 245, 246, 286
- O**
 optimal time allocation 291–3
- P**
 photon counting 142–4, 284, 301–2, 304,
 308–10, 311, 315
 photon detection efficiency (PDE) 14–15
- Pockels effect 517
 polytetrafluoroethylene (PTFE) 81–3
 power meters 426
- R**
 reflex anode 340
- S**
 scintillators 52–3, 83–4, 86–90, 93–5
 afterpulses 180–3, 197, 267, 315, 433–4
 attenuation length 80–1
 correlated signals 197–200
 counters 105–6
 early pulses 437, 438
 efficiency 101–7, 269
 fibres 110–12
 housings 462–3
 inorganic 86, 89
 large area 95, 101–5, 114
 late pulses 432, 435
 liquids 90
 plastic 86
 prepulses 432, 435, 436, 438
 tiles 112–13
 Tyvek 83
 uniformity of response 88–9, 107–8
 signal induced background 432–53
 small area 103
 spectrometers 130
 silicon PIN diodes 12
 single channel devices 237–9
 single-electron response (SER) 7, 136,
 153, 166–7, 208–9
 time response 351–2
 single-photoelectron distributions 592–6
 single-photon sources 130–1
 Snell's law 37–8
 solid light guides 98–9
 Sommer, H A 3
 statistics 4–7, 20, 129–200, 369–76
 arrival times 175–80
 binomial distribution 129, 131–2,
 133, 136
 cascaded processes 153–61
 central moments 140–1
 correlated signals 197–200
 expectation values 133–5
 exponential distribution 136
 factorial moments 141–2, 143, 144
 folding 137–9, 151–3, 343
 generating functions 145–6
 mean and variance 132–6
- moment generating functions 149–51
 moments 139–42, 143
 multiplier cascade processes 155–61
 normal distribution 132, 133
 poisson distribution 129, 131–2,
 133, 135
 Polya distribution 222–4
 probability distributions 131–2
 probability generating functions 145–6,
 151–3
 variance 132–6
- T**
 temperature effects 65–9, 218, 267–9,
 314, 467–70
 threshold levels 300–1, 316–17, 351,
 361–2
 time intervals 184–7, 298, 343–5
 total internal reflection (TIR) 37–9
 transients 502–3, 534–5
 transit times 323, 331, 337–40, 353–6,
 370–3
 transmissivity 47
 trap detectors 35–7
- U**
 undersized signals/pulses 219–25, 267
 uniformity of response 107–8
- V**
 voltage dividers 472–514
 analysis of resistive dividers 478
 Cockcroft Walton dividers 497–9
 decoupling capacitors 499–502
 dynode output 507–9
 fully active 488–95
 gain adjustment 511
 gain linearity 483, 480–4, 485
 including Zener diodes 475–7, 484–8
 partially active 496–8
 pin configurations 501, 506–7
 power supply polarity and ripple 504
 pulsed operation 499
 resistive 473–5, 477
 shorting dynodes 509–10
 single cable dividers 510
 switch-on transients 502–3
 wiring practice 512–14
- Z**
 zener diodes 488