

**COINCIDENCE ANALYSIS  
OF  
GRAVITATIONAL WAVE DATA**

**C.A. DICKSON**

A thesis submitted to the University of Wales  
for the degree of Doctor of Philosophy  
September 1993

# Declaration

I declare that this work has not already been accepted in any substance for any degree, and is not being currently submitted in candidature for any degree.

---

(Candidate)

Except where otherwise stated, this work is wholly the result of the candidate's own investigation. Suitable credit is given to joint work with colleagues, and to work of others throughout the thesis.

---

(Candidate)

---

(Supervisor)

# Acknowledgements

I would like to thank my supervisor, Professor Bernard F. Schutz, for his guidance and countless helpful suggestions.

I would also like to acknowledge the help and support of all at the Department of Physics & Astronomy, University of Wales College of Cardiff; particularly Gabrielle Allen, Tim Barnett, Simon Chapman, Rob Davies, Dean Johnson, Gareth Jones, John Bryn Jones, David Nicholson, Justin Shuttleworth and John Watkins. Apologies to anyone whom I have missed.

Thanks are also due to many outside my institution who have assisted me, especially David Robertson at the University of Glasgow and Albrecht Rüdiger at the Max-Planck-Institut für Quantenoptik, Garching.

I acknowledge financial support from the S.E.R.C.; studentship number 8930551X.

*To my Father and Mother*

## Abstract

The work presented herein falls into three parts. Part I reassesses the claims recently made by the Rome-Turin-Maryland (RTM) collaboration, that about the time of Supernova SN1987A, there were unusual correlations observed between four particle detectors and two room-temperature bar gravitational wave detectors. These correlations were claimed to have chance probability of as low as  $\sim 10^{-6}$ . By evaluation of RTM's *a posteriori* adjustment of many free parameters, I revise the probability estimates up to between  $\sim 10^{-3}$  and the level of chance. I conclude, in contradiction to RTM, that the correlations are more likely due to chance fluctuations in the data than to a new physical effect.

Part II is a short, mainly discursive, section. Here, I state many lessons which can be learned from RTM's analysis, with particular relevance to the coincidence analysis which I perform in Part III.

In Part III, I perform the first coincidence analysis of data taken from interferometric gravitational wave detectors, the data coming from a coincident experiment lasting 100 hours (the *100 Hour Data Run*) in March 1989, between the prototype detectors at the University of Glasgow and the Max-Planck-Institut für Quantenoptik, Garching. In particular, I present the first working program for the coincidence analysis of data taken from two interferometric detectors. I devise efficient methods for vetoing untrustworthy data, including the *h-veto*, which removes coincidences which have measured amplitudes differing by more than a predetermined amount in probability space.

After applying these vetoes, I show that there were no highly improbable coincidences during the experiment. I place the first experimental limits on 10 kHz broadband gravitational waves: no coincidences were seen above  $h = (6.8 \pm 1.3) \times 10^{-16}$  during the experiment. I also present a way in which this limit could be improved for future similar experiments. Finally, I list the lessons learned from my coincidence analysis, for the interest of experimenters and data analysts in the field.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Astrophysics of gravitational waves . . . . .	2
1.1.1	Gravitational waves in general relativity . . . . .	2
1.1.2	Astrophysical sources of gravitational waves . . . . .	3
1.2	Detectors of gravitational waves . . . . .	5
1.3	State of affairs in gravitational wave research . . . . .	8
1.4	Topics covered in this thesis . . . . .	10
<b>I</b>	<b>Gravitational Wave – Neutrino Correlations</b>	<b>13</b>
<b>2</b>	<b>Reassessment of the Reported Correlations between Gravitational Waves and Neutrinos Associated with SN1987A</b>	<b>14</b>
2.1	Introduction . . . . .	15
2.2	The Gravitational Wave and Neutrino Observations . . . . .	20
2.2.1	Particle observations . . . . .	20
2.2.2	Gravitational wave detectors . . . . .	21
2.3	Summary of the Main RTM Analysis Methods . . . . .	22
2.3.1	The main RTM analysis methods . . . . .	22
2.4	Monte Carlo simulations . . . . .	26
2.4.1	Computer model . . . . .	26
2.4.2	Properties of the pseudo-random data . . . . .	26
2.4.3	Method of analysis of the pseudo-random data . . . . .	27
2.4.4	Results . . . . .	35
2.5	Reassessment of RTM correlations . . . . .	43

2.5.1	Reassessment of Mt. Blanc–gravitational wave coincidences . . . . .	44
2.5.2	Reassessment of KII–gravitational wave coincidences . . . . .	51
2.5.3	Reassessment of gravitational wave–IMB correlations . . . . .	56
2.6	Conclusions . . . . .	57
2.7	Appendix: Review of the RTM analyses . . . . .	59
2.7.1	The gravitational wave–Mt. Blanc coincidences . . . . .	59
2.7.2	The Rome–Maryland gravitational wave coincidences . . . . .	62
2.7.3	The gravitational wave–KII–Mt. Blanc coincidences . . . . .	64
2.7.4	The gravitational wave–IMB–Mt. Blanc coincidences . . . . .	67
<b>3</b>	<b>Reassessment of the Reported Mont Blanc–Baksan Neutrino Coincidences</b>	<b>70</b>
3.1	Chudakov’s Assessment of Mont Blanc–Baksan Coincidences . . . . .	71
3.2	Reassessment of Mont Blanc–Baksan Coincidences . . . . .	71
3.2.1	Choice of dataset . . . . .	72
3.2.2	Clock correction . . . . .	72
3.2.3	Delay window . . . . .	74
3.3	Corrected probability . . . . .	74
3.4	Conclusions for Mont Blanc–Baksan Correlations . . . . .	75
<b>II</b>		<b>76</b>
<b>4</b>	<b>Lessons Learned for Coincidence Analysis and Pioneering Research</b>	
	<b>at Low S/N</b>	<b>77</b>
4.1	Post mortem on the RTM analyses . . . . .	77
4.2	Models: working with and without them . . . . .	80
4.2.1	Analysis of noise . . . . .	80
4.2.2	Working models and the null hypothesis . . . . .	82
4.2.3	When to change the model . . . . .	83
4.3	Low S/N analysis in physical sciences . . . . .	86
4.4	Lessons learned for gravitational wave coincidence analysis . . . . .	88

<b>III</b>	<b>Coincidence Analysis of the 100 Hour Experiment</b>	<b>90</b>
<b>5</b>	<b>Background to the Coincidence Analysis of the 100 Hour Experiment</b>	<b>91</b>
5.1	Background to the experiment . . . . .	92
5.1.1	Objectives of the 100 Hour Experiment . . . . .	92
5.1.2	Experimental details . . . . .	92
5.1.3	Data storage and manipulation . . . . .	94
5.2	Individual analysis of the Glasgow and Garching data streams . . . . .	95
5.2.1	Individual analysis of the Glasgow data stream . . . . .	96
5.2.2	Individual analysis of the Garching data stream . . . . .	98
<b>6</b>	<b>Coincidence Methodology</b>	<b>102</b>
6.1	Analysis strategy . . . . .	102
6.1.1	General strategy . . . . .	102
6.1.2	Assessment of significance of results . . . . .	107
6.1.3	Vetoing poor data . . . . .	108
6.1.4	Brief preview of analysis of filtered output . . . . .	108
6.2	Astrophysical model . . . . .	110
6.2.1	Expected sources . . . . .	110
6.2.2	Effects of detector geometry . . . . .	111
6.3	Expected laboratory-based effects . . . . .	115
6.4	Coincidence program . . . . .	117
6.4.1	My coincidence program . . . . .	117
6.4.2	Nicholson's adaptation of the coincidence program . . . . .	118
<b>7</b>	<b>Vetoos</b>	<b>120</b>
7.1	Data available for vetoing . . . . .	121
7.2	Primary vetoos . . . . .	123
7.2.1	Method of primary veto selection . . . . .	123
7.2.2	Choice of primary vetoos . . . . .	125
7.2.3	Initial effects of these choices . . . . .	134
7.3	Secondary vetoos . . . . .	138
7.3.1	The Glasgow Gaussian Parameter . . . . .	138



7.3.2	The Munich Gaussian Parameter . . . . .	143
7.3.3	Duration discrepancy . . . . .	144
7.3.4	Amplitude discrepancy . . . . .	146
7.3.5	Sensitivity veto (sigma veto) . . . . .	155
7.3.6	Event sign veto . . . . .	156
7.4	Interrelations of the vetoes . . . . .	156
7.4.1	Interrelations of Glasgow vetoes . . . . .	157
7.4.2	Interrelations of Munich vetoes . . . . .	160
<b>8</b>	<b>Results of the Coincidence Analysis</b>	<b>165</b>
8.1	Coincidence list . . . . .	165
8.2	Probability analysis . . . . .	172
8.2.1	Method of probability calculation . . . . .	172
8.2.2	Results of probability calculation . . . . .	175
8.3	Upper limit on $h$ . . . . .	177
<b>9</b>	<b>Conclusions of the Coincidence Analysis</b>	<b>180</b>
9.1	Astrophysical conclusions . . . . .	180
9.2	Experimental conclusions . . . . .	181
	<b>Appendix A: Listing of Coincidence Program (Prototype Version)</b>	<b>183</b>
	<b>Appendix B: Comments on the 100 Hour Experiment</b>	<b>204</b>
	<b>References</b>	<b>210</b>

# List of Figures

1.1	Basic design of a laser interferometric gravitational wave detector . . . . .	6
2.1	Distribution of the pseudo-random number generator . . . . .	27
2.2	First indications of the Mt. Blanc correlations . . . . .	30
2.3	Searching for good time delays using the net-excitation method applied to the data set of Fig. 2.2, using Rome data only . . . . .	31
2.4	Searching for good time delays using the net-excitation method applied to the data set of Fig. 2.2, using both Rome and Maryland data . . . . .	32
2.5	Histogram of number of threshold-coincidences against delay time for data set 55 of the first simulation run. . . . .	36
2.6	Result of the net-excitation analysis of simulation set 55 (a) compared to the RTM analysis of the real “neutrino” and gravitational wave data (b) . . . . .	38
2.7	Delay histogram for data set 327 of the second simulation (a) compared to the RTM histogram of the real data (b) . . . . .	40
2.8	Net-excitation analysis of set 327 . . . . .	41
2.9	Relative frequency distribution of the values of the parameter $q$ in our second simulation run . . . . .	42
2.10	Logarithm of the previous figure, showing a nearly exponential distribution . . . . .	42
2.11	Frequency distribution of $q$ for small $q$ . . . . .	43
2.12	The energy distribution of the IMB particles during the period in question . . . . .	68
5.1	Glasgow noise distribution while the detector cavities are locked . . . . .	97
5.2	Glasgow noise distribution while the detector cavities are out of lock . . . . .	98
5.3	Glasgow noise spectrum . . . . .	99
5.4	Garching noise spectrum . . . . .	100

6.1	Example gravitational wave received in two detectors, where the peaks of the two output events are more than the wave travel time apart . . . . .	104
6.2	Coincidences produced when the progenitor events are (a) far apart, and (b) close together in time, compared to the coincidence window . . . . .	106
6.3	Relative geometry of two detectors observing a distant source . . . . .	112
7.1	Glasgow Event Primary Error Point Signals (coincidence data 5h–93h) . . . .	127
7.2	Glasgow Group Primary Error Point Signals (coincidence data 5h–93h) . . .	127
7.3	Glasgow Event Microphone Signals (coincidence data 5h–93h) . . . . .	128
7.4	Glasgow Group Microphone Signals (coincidence data 5h–93h) . . . . .	128
7.5	Glasgow Group max Secondary Visibility Signals (coincidence data 5h–93h) .	129
7.6	Glasgow Event Seismic Signals (coincidence data 5h–93h) . . . . .	130
7.7	Glasgow Group Seismic Signals (coincidence data 5h–93h) . . . . .	130
7.8	Munich Event Table Seismometer Signals (coincidence data 5h–93h) . . . . .	131
7.9	Munich Group Table Seismometer Signals (coincidence data 5h–93h) . . . . .	131
7.10	Munich Event Suspension Seismometer Signals (coincidence data 5h–93h) . .	132
7.11	Munich Group Suspension Seismometer Signals (coincidence data 5h–93h) . .	132
7.12	Munich Event Low Frequency Signals (coincidence data 5h–93h) . . . . .	133
7.13	Munich Group Low Frequency Signals (coincidence data 5h–93h) . . . . .	133
7.14	Munich Group Suspension Seismometer Signals (coincidence data 5h–46h) . .	135
7.15	Munich Group Suspension Seismometer Signals (coincidence data 46h–93h) .	135
7.16	Distribution of Gaussian Parameters from 14000 simulated Glasgow groups .	139
7.17	Sample Glasgow Output . . . . .	140
7.18	Distribution of Glasgow Gaussian Parameters (coincidence data 5h–93h) . . .	141
7.19	Distribution of Munich Gaussian Parameters (coincidence data 5h–93h) . . .	143
7.20	Example gravitational wave duration above the threshold of two detectors . .	145
7.21	Area of integration over which we veto a coincidence with discrepant gravitational-wave amplitude . . . . .	150
7.22	Munich group data (57h 20m to 62h 34m 45s) . . . . .	163
8.1	Numbers of coincidences in each hour (5h–93h, no vetoes) . . . . .	167
8.2	Numbers of coincidences at each delay between observations (5h–93h, no vetoes)	167
8.3	Scatter diagram of the signal-to-noise of the coincidences (5h–93h, no vetoes)	169

8.4	Scatter diagram of the signal-to-noise of the coincidences (5h–93h, all vetoes except $h$ -veto) . . . . .	170
8.5	Scatter diagram of the signal-to-noise of the coincidences (5h–93h, all vetoes)	171
8.6	Simultaneous events at the two detectors, with probability $p_g p_m$ . . . . .	173
8.7	Distribution of signal-to-noise of Glasgow events (event list, 5h–93h, all Glasgow vetoes) . . . . .	176
8.8	Distribution of signal-to-noise of Munich events (event list, 5h–93h, all Munich vetoes) . . . . .	176
8.9	Distribution of empirical probabilities of coincidences (coincidence list, 5h–93h, all vetoes) . . . . .	177
8.10	Distribution of peak $h$ values for coincidences, averaged over the two detectors (5h–93h, all vetoes) . . . . .	179

# List of Tables

1.1	Collaborations to build long interferometric gravitational wave detectors . . .	9
1.2	Existing gravitational wave bar detectors . . . . .	10
5.1	Specifications of the Glasgow and Garching prototypes . . . . .	93
7.1	Data recorded at Glasgow during 100 hour experiment . . . . .	122
7.2	Data recorded at Munich during 100 hour experiment . . . . .	123
7.3	Values of vetoes adopted from histogram analysis. . . . .	126
7.4	Numbers of coincidence events removed by vetoes . . . . .	136
7.5	Effect on Gaussian Parameter on introducing an artificial event into the group	142
7.6	Overlap of Glasgow vetoes . . . . .	158
7.7	Overlap of Munich vetoes . . . . .	161
9.1	Upper limit on $h$ for various vetoing systems . . . . .	181

# Chapter 1

## Introduction

Gravitational waves were theoretically predicted by Einstein in 1916, as a consequence of his general theory of relativity. Their existence is the most important remaining test of general relativity, and observation of gravitational waves will reveal much new astrophysics which will be interesting in its own right.

However, gravitational waves have not yet been convincingly detected, at the time of writing. Gravitational waves are very weak as a phenomenon; and at present, neither the detectors nor the data analysis systems are adequate to achieve reasonable observation rates. The solution of this detection problem and, later, the establishment of an observational science of gravitational wave astronomy, depends on attacking the problem from these two directions. On the one hand, one must build detectors which are sensitive enough. On the other hand, one must devise analysis methods and software to find the signals which may be there and, for the most part, to do this automatically and in real time.

With the next generation of detectors now being planned, we expect the increase in sensitivity obtained to facilitate the detection of gravitational waves by the end of the millenium. The overall system of data analysis also requires such an improvement. This thesis concerns itself with some of the remaining unsolved problems in the data analysis of gravitational waves. I hope it will contribute to the important first detection, when it happens; and, later, to the establishment of an observational science of gravitational wave astronomy.

## 1.1 Astrophysics of gravitational waves

### 1.1.1 Gravitational waves in general relativity

The field equations of general relativity, as derived by Einstein, are

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = kT_{\mu\nu}; \quad (1.1)$$

where  $R_{\mu\nu}$  is the Ricci tensor,  $R = R^\mu{}_\mu = g^{\mu\nu}R_{\mu\nu}$  is the trace of the Ricci tensor (the *Ricci scalar*),  $g_{\mu\nu}$  is the metric tensor,  $T_{\mu\nu}$  is the stress-energy tensor, and  $k$  is a constant. I use the following conventions:

- Strictly,  $g_{\mu\nu}$  is not the metric tensor, but the  $(\mu, \nu)$  component of the metric tensor, normally denoted  $g$ ; but I shall continue to use this lazy terminology. This also applies to the other “tensor” terms in the equation.
- The Greek indices  $\mu, \nu$ , etc. take values 0, 1, 2, 3. In flat space, I shall interpret these as the usual cartesian coordinates of special relativity, i.e.  $t, x, y, z$ , respectively.
- I use repetition of indices to indicate summation over the indices in question (the *summation convention*).

See, e.g. Schutz (1985) for more details. As is fairly standard for this calculation, I have ignored the cosmological constant term.

Consider the vacuum solution ( $T_{\mu\nu} = 0$ ), and the weak field approximation, i.e.,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (1.2)$$

where  $\eta_{\mu\nu}$  is the Galilean metric of flat spacetime, and  $h_{\mu\nu}$  is a small perturbation, i.e.  $|h_{\mu\nu}| \ll 1$ . Now if we linearise and adopt the Lorentz gauge (see Schutz 1985), Eq. 1.1 becomes

$$\square h_{\mu\nu} \equiv \left( -\frac{\partial^2}{\partial t^2} + \nabla^2 \right) h_{\mu\nu} = 0. \quad (1.3)$$

This is d'Alembert's equation. The most simple solution is

$$h_{\mu\nu} = \text{Re} \left[ A_{\mu\nu} e^{2\pi i(t-z/c)} \right], \quad (1.4)$$

that is, the equation of a three dimensional wave propagating through spacetime in the  $z$  direction, at the speed of light,  $c$ . This is a gravitational wave. We shall return to their effect on matter, and how they are to be detected, in a moment. Firstly, a quick review of expected astrophysical sources of gravitational waves.

### 1.1.2 Astrophysical sources of gravitational waves

Gravitational waves are produced by all matter which is moving with a non-zero quadrupole moment. How much energy is released, and whether the waves are observable from sources at astrophysical distances, is another question. Even the most energetic sources predicted will be very difficult to observe with Earth-based detectors, because gravitational waves generally couple very weakly with matter. At present, there are four main expected sources of gravitational waves which we expect will be observable in ground-based detectors. These are the following (for more details, see, e.g. Thorne 1987).

#### Stellar collapse

At the end of its life, a star will suffer one or more collapses, because its radiation and gas pressures can no longer sustain the star against its own inward gravitational pull. As the core of the star collapses, it can give off gravitational radiation. Although only one supernova every hundred years or so is expected in a galaxy of our size, there may be many more collapses which are *electromagnetically-quiet*, i.e. do not have dramatic supernova-type optical or electromagnetic displays, or which are hidden in or behind dense gaseous clouds. Current guesses at event rates are of the order of one collapse per thirty years in our galaxy.

Of course, if one can observe out to more distant galaxies and other clusters of galaxies, the event rate will go up in proportion to the volume of space observed, i.e. to the cube of the distance out to which one can observe these phenomena. The prototype detectors working in Glasgow and Garching (see Section 1.2), and whose data I analyse in coincidence (see Part III), could only barely detect a nearby collapse event in our galaxy; while the next generation of long interferometric gravitational wave observatories (such as LIGO and VIRGO), with their much increased sensitivities, are expected to see collapse sources out to the Virgo Cluster, with an expected event rate of around several hundred per year (Hough *et al.* 1989).

#### Coalescing compact binary stars

All binary stars, due to the non-zero quadrupole moment of their orbit, are gradually losing energy in the form of gravitational waves. This will cause the orbit to decay<sup>1</sup>, such that even-

---

<sup>1</sup>as is seen directly in the pulsar PSR 1913+16, whose orbital period is decreasing in exactly the way predicted by general relativity, to the observational limit of 1% of the effect (the effect being that the rate



tually the stars will coalesce. Calculations for compact binaries (see e.g. Schutz 1986; Thorne 1987 and references therein) show that most of the observable gravitational wave energy is given off in the last few seconds, when the orbital frequency reaches 100 Hz and above. (The close orbit of gaseous stars is much more difficult to simulate, due to hydrodynamics and tidal effects.)

Although the expected event rate for coalescing binaries in our galaxy is very low, compared to collapse event rates, the fact that one can fairly accurately predict the waveform of binary coalescence enables one to employ data analysis techniques, such as matched filtering, to improve the signal-to-noise ratio for a given detection; and hence to see objects much further away than would otherwise be the case. Hence, the observed event rate may be comparable to that of collapse events; but this is model-dependent (see Phinney 1991).

### Continuous wave sources

Any rapidly rotating object will emit gravitational waves, if it has a non-zero quadrupole moment. Rapidly rotating neutron stars are the most famous candidate sources of this type of gravitational radiation. In order for them to emit, however, they must have some kind of non-axisymmetry (the axis in question being the rotation axis) due to either “geophysical” deformations (such as mountains) or large scale eccentricity of shape (e.g. caused by mechanical instabilities or magnetic effects). In this case, the expected wave form will be a sine wave, or several sine waves superimposed in the case of several non-axisymmetric imperfections in the shape of the body.

The expected amplitude of waves from such a source will be much lower again than for coalescing binaries; but the very long period of possible observation (up to years, interruptions being unimportant so long as the phase is preserved), coupled with our knowledge of the waveform, will enable observations of many sources in our galaxy with the long interferometers, operating in broadband. The main unsolved problem here in the data analysis is the inversion of the Doppler motion of the Earth, which is unknown *a priori* when the rotation frequency and location of the source on the sky are unknown. Some attempts have been made to solve this problem: see e.g. Schutz (1991).

---

of change of the orbital period,  $\dot{P}_b = -2.43 \times 10^{-12} \text{ s s}^{-1}$ ). See Taylor & Weisberg 1989. This is almost conclusive observational evidence that gravitational waves exist, and are emitted at energies predicted by general relativity.

### Stochastic background

Finally, it is expected that there will be other continuous sources which, although observable in amplitude terms, have unmodelled waveforms. These form the stochastic background of gravitational waves. Theoreticians have predicted that cosmic strings, if they exist, will be an observable source of gravitational waves. Furthermore, the Big Bang itself should be the source of an observable “echo” of gravitational waves (analogous to the 3 K cosmological microwave background). These waves are expected to be seen by the cross-correlation of the outputs of two or more detectors operating in coincidence over long periods.

## 1.2 Detectors of gravitational waves

Return now to Eq. 1.4. By imposing additional gauge conditions, and choosing the observers coordinate system, we can force the only non-zero components of  $A_{\mu\nu}$  to be (see e.g. Schutz 1985),

$$A_{11} = -A_{22} \equiv A_+, \quad (1.5)$$

and

$$A_{12} = A_{21} \equiv A_x. \quad (1.6)$$

Hence, from the geodesic differential equation (see Schutz 1985), one can show that if two particles are separated by  $\epsilon$  in the  $x$  direction, the distance between them changes as the gravitational wave passes, such that

$$h_{xx} = -h_{yy} = \frac{2\delta\epsilon}{\epsilon}, \quad (1.7)$$

where  $\delta\epsilon$  is the change in distance.

Thus, gravitational waves interact with matter by causing motion of particles in the plane perpendicular to that of wave propagation. The amplitudes of oscillation are dependent firstly on the amplitudes of the two *polarization* components,  $A_+$  and  $A_x$ , of the wave; and secondly on the distance between the bodies. The motion is such that, for example, if the wave has only the “+” polarization component, the space between two bodies in the  $x$  direction contracts as the space in the  $y$  direction expands, and vice versa.

This motion is the principle behind the two main types of gravitational wave detectors. These are as follows.

### Laser interferometric gravitational wave detectors

The most common design for an interferometric gravitational wave detector is the Michelson interferometer. This exploits the motion induced in the two dimensions perpendicular to the direction of wave propagation. See Fig. 1.1.

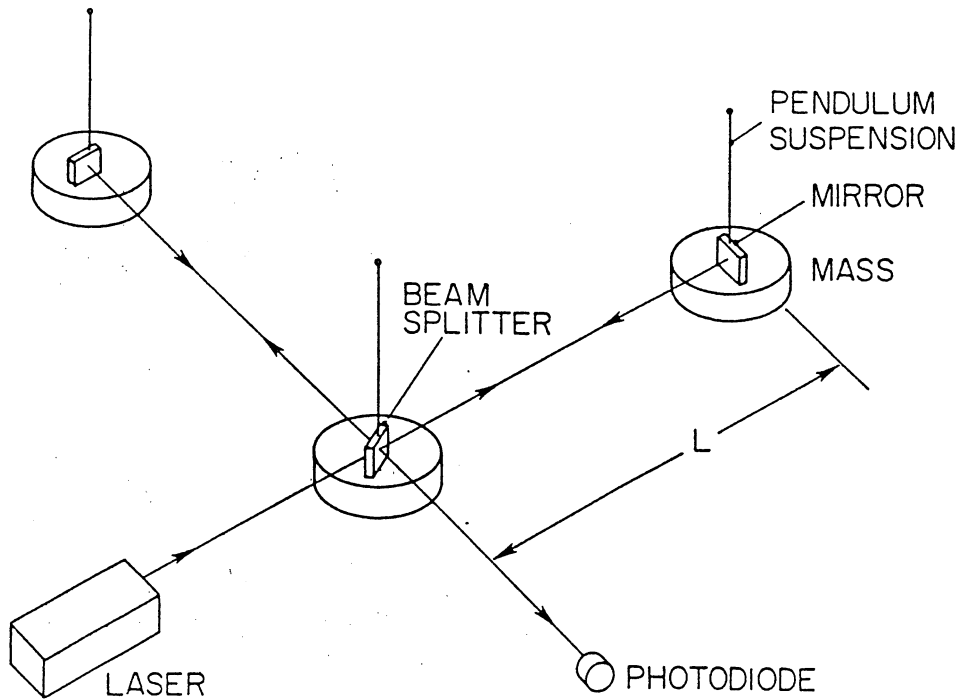


Figure 1.1: Basic design of a laser interferometric gravitational wave detector. Taken from Hough *et al.* (1986).

The mirrors at the ends of the arms reflect the laser light back to the beam splitter, where an interference pattern is observed. If a gravitational wave passes, the mirrors swing and the interference pattern changes. The effect is, however, very small. For a typical burst source in the Virgo cluster, with  $h \sim 10^{-22}$ , and with an effective arm length (many bounces of laser light are used in the arms of the interferometer, to increase the effective path distance between the test masses and the beam splitter) of  $10^5$  m, Eq. 1.7 gives the measured change in position of the masses as  $10^{-17}$  m — smaller than the diameter of a proton.

Interferometric detectors are further subdivided into Fabry-Perot (resonant cavity) and delay line interferometers. I shall not go into more detail on these two types of interferometer: see, e.g., Hough *et al.* (1989).

At present, there are four working prototype laser interferometers in the world. These are situated at: the University of Glasgow; the Max-Planck-Institut für Quantenoptik, Garching, Germany; the California Institute of Technology; and at ISAS, Tokyo. The Glasgow prototype, for example, is of the Fabry-Perot design, while the Garching prototype is of the delay-line design.

### Resonant bar detectors

Resonant bar detectors are slightly simpler in concept than are laser interferometers. A bar detector is a large cylinder of metal, usually aluminium alloy. If a gravitational wave passes the bar, at the appropriate frequency (the resonant frequency of the bar), it will excite the bar and cause it to “ring” at that frequency. A transducer converts this excitation to an output voltage. Again, the motions measured are very small; and thermal noise is the dominant noise source. Therefore modern bars are cooled down to cryogenic temperatures. About the time of the supernova SN1987A, typical bars functioned at room temperatures; while the new NAUTILUS ultracryogenic bar in Rome has recently been tested at below 100 mK (Astone *et al.* 1993).

The resonant frequencies of most bar detectors are around 1 kHz, about the expected frequency of the gravitational radiation emitted from collapse sources. This is also around the expected frequency of peak amplitude of gravitational wave emission from coalescing compact binaries.

### Interferometers versus bars

For these two main types of detector, each has advantages over the other. The interferometer is broadband, i.e. is relatively sensitive to gravitational waves from frequencies of around 50 Hz right up to their sampling frequency, which may be 10 or 20 kHz. The bar is only sensitive in a narrow band (of the order of 2 Hz — Astone *et al.* 1993) about its resonant frequency. The interferometer doesn’t need cooling as does the bar. Furthermore, the bar design seems to have a built-in sensitivity limit, given by the thermal quantum motions of the particles, at a sensitivity of around  $h \sim 10^{-21}$  (Astone *et al.* 1993), which seems hard to beat no matter how low is the temperature. Unfortunately, this would seem to preclude the detection of bursts outside our own galaxy. The interferometer doesn’t seem to suffer from this problem, in principle at least.

On the other hand, the bar is much cheaper and smaller than an interferometer of “comparable sensitivity” (if one can compare the sensitivities of a broadband and a narrowband detector). The interferometer requires expensive and bulky vacuum tubes, which the bar does not, and the interferometer arms would need to be hundreds of metres or even several kilometres long, in order to beat the bar in sensitivity at the bar’s own resonance frequency (and to have reasonable event rates for detecting bursts and binary coalescences). The interferometer also requires expensive and advanced laser and mirror technology to operate at optimum sensitivity with arm lengths of the order of 3 km; which is the length scale on which LIGO are currently building and VIRGO are planning to build.

The basic choice between detectors of the interferometer or bar type is between broadband sensitivity and inexpensiveness; which is a matter of science versus funding. Of course, both types will be observing in the near future. The information they will provide will be complementary, and they may even be directly used together in coincidence (see e.g. Astone, Lobo & Schutz 1993).

### 1.3 State of affairs in gravitational wave research

Experimental gravitational wave astronomy as such began in the late 1950’s with J. Weber, who constructed the first bar detector (Weber 1960). Of course, the sensitivity of the bar was far removed from what we expect are the sensitivity levels needed to observe gravitational waves from a distant source. Since then, many improvements have been made in bar technology (see e.g. Astone *et al.* 1993).

Prototype interferometers have also been built and run in the last twenty or so years. Again, great improvements have been made of late (see e.g. Newton 1993).

However, as I have said, there has not yet been a convincing claim for the detection of gravitational waves. There have been claims, but none of them has stood up to close scrutiny. We shall return to one such claim in Chapter 2, which at least involves bar detectors, even though the experimenters have not postulated a coherent model to explain what they see, nor even whether they believe that gravitational waves are actually involved.

Experimental gravitational waves, at least up to the time of writing, has been largely concerned firstly with setting better and better upper limits on sources; and secondly, in the case of interferometers, with improving the sensitivity of prototype detectors at various

Table 1.1: Collaborations to build long interferometric gravitational wave detectors

Collaboration	Institutions	Arm Length	Timescale <sup>2</sup>
LIGO	Caltech & MIT	$2 \times 4$ km	1998
VIRGO	INFN Pisa & CNRS Orsay/Paris	3 km	?
GEO	U. Glasgow, MPQ Garching, & Cardiff U.	3 km ?	?
AIGO + collaborators	U.W. Australia, ANU, CSIRO Adelaide; IUCCA & CAT (India)	3 km	?
TENKO	ISAS (Tokyo)	100 m	1995

frequencies, and removing various noise sources in the detectors.

The task facing both the experimenters and data analysts is a huge one: much time and effort is being invested in solving all the problems associated with (1) the detection of the first gravitational wave, and (2) turning this nascent branch of astrophysics into a working observational science. There has been much collaborative effort amongst experimental teams, and between experimental teams and data analysis specialists: my own coincidence analysis of the joint 100 Hour Experiment is one of the fruits of the collaboration between the experimental groups at the University of Glasgow, Scotland and the Max-Planck-Institut für Quantenoptik in Garching, Munich, Germany, and the data analysis team at University of Wales College of Cardiff.

Long interferometers are very expensive and complicated to construct. Consequently, all but one of the current plans to build such detectors involve some sort of collaboration across institutions. At present, the main collaborative projects in the field are as shown in Table 1.1 (this is intended only as a rough guide, and in no way should be taken as a definitive statement on the subject).

The subject of collaborations has unavoidably become a sensitive political issue, so I will not dwell on this here. Note also that there are at least two proposed projects to put interferometers in space, namely LISA and SAGGITARIUS, but the funding and institutions involved in these are even more complicated.

The *status quo* in bar technology is simpler. The locations of the main bar detectors,

---

<sup>2</sup>Projected completion of building and preliminary operation.

along with a brief description of the types and temperatures of the bars, are given in Table 1.2.

Table 1.2: Existing gravitational wave bar detectors. (Based on Schutz 1989)

Institution	Room-temperature	Torsion	Cryogenic	Ultracryogenic
Stanford U.			✓	✓
Louisiana State U.			✓	
U. of Maryland	✓		✓	
U. of Rome	✓		✓	✓
U.W. Australia			✓	
Moscow State U.	✓			
Tokyo U.		✓		
Guangzhou & Beijing	✓			

The future for experimental gravitational wave research looks increasingly bright, with several long interferometers now at the planning stages, and with bars becoming more and more sensitive as they are cooled to lower temperatures. The data analysis problem, however, is a huge and underrated task in itself. It is with part of this problem that my thesis is concerned. The main topics I shall cover are as follows in the next section.

## 1.4 Topics covered in this thesis

### Part I

The claims by the RTM (Rome-Turin-Maryland collaboration) groups, that there were strong correlations between various gravitational wave and particle detectors around the time of the supernova SN1987A in the Large Magellanic Cloud have stood unchallenged in the literature for some time. Indeed, RTM are still producing papers concerning this matter. However, many researchers believe that these correlations are more likely due to an artefact of RTM's data analysis than due to real signals of some sort, whether gravitational waves, neutrinos, some new particle or interaction, or a combination of these.

My doctoral supervisor and I have written a paper (Dickson & Schutz 1992) which performs a detailed reassessment on their analysis, complete with an attempt to account for every a *posteriori* choice or change of analysis parameters which we believe RTM used in

order to optimise the correlations which they see. Our paper rebuts RTM's claims for very unlikely correlations, and we find that the correlations are very much more likely to have arisen by chance than RTM claim. Hence, we believe, there is no compelling evidence to conclude that gravitational waves or anything else have been seen in these data. A version of this paper is included verbatim as my Chapter 2. Although much collaboration with my supervisor was necessary in the writing of this paper, most of the original ideas are my own.

Following on from this is what could be described as an addendum to the paper contained in Chapter 2. This looks at some more claimed coincidences, similar to those in Chapter 2 and apparently occurring at the same time, between two neutrino detectors. This was not included in the original paper. It was written by me with very little discussion with my supervisor. The remainder of the thesis is exclusively my own work.

## Part II

In Chapter 4, which forms the whole of Part II, I look at the lessons which can be learned from RTM's analysis of the gravitational wave and particle data. I find that they have a completely flawed approach to the analysis of these data, and draw various conclusions concerning the way I believe analysis of such data sets should and should not be performed. This generalises to other problems in physical sciences where the data have low signal-to-noise ratio, and where one may want to consider unmodelled or unexpected objects or even new physical processes as being candidate sources of signals in the data.

I use these lessons learned to make a short list of recommendations for future such analysis of gravitational wave data.

## Part III

Bearing my own recommendations in mind, I then perform the first coincidence analysis of laser-interferometric data. Those data were taken from a coincidence experiment lasting 100 hours (the 100 Hour Experiment, or 100 Hour Data Run), in March 1989, between the prototype interferometers in Glasgow and Garching, to which I have already referred.

In Chapter 5 I review the background to the experiment, so the reader is familiar with the necessary details before I proceed.

Chapter 6 describes my methodology for the coincidence analysis, including my method for estimating the probability of any coincidences occurring by chance during the experiment.



I choose to concentrate on short duration collapse and burst sources for the coincidence analysis.

I tackle the large subject of *veto*s in Chapter 7. This involves examining the various housekeeping data streams (seismometers, microphone signals, etc.) recorded in the laboratories of the two detectors, and using them to remove, or *veto*, events or coincidences which I deem untrustworthy. I also consider ways of removing coincidences which are untrustworthy for reasons other than that the housekeeping data were exhibiting unusual behaviour at certain times.

In Chapter 8 I give the results of my coincidence analysis. I set an upper limit on broadband bursts during the 100 Hour Experiment, and show that no highly improbable coincidences have occurred during the experiment.

My conclusions for the coincidence analysis are given in Chapter 9. I also list there the main achievements of my research in the coincidence analysis and in the writing of Part III.

At the end of my thesis I include two Appendices. Appendix A is a listing of my coincidence program, which was the first working program written to search for threshold-crossing coincidences in interferometric gravitational wave data. Appendix B is a list of my recommendations for future experiments, based on my experience with the coincidence analysis of the 100 Hour data. These recommendations have relevance to future experiments with other, larger interferometers, as well as to future experiments with these prototype detectors. They also pertain both to experiment and to data analysis, as these fields overlap considerably in practise.

## Part I

# Gravitational Wave – Neutrino Correlations

## Chapter 2

# Reassessment of the Reported Correlations between Gravitational Waves and Neutrinos Associated with SN1987A<sup>1</sup>

### Abstract

Correlations of considerable apparent significance have been reported between data taken by two bar-type gravitational wave detectors and particle events recorded in the Mt. Blanc, Kamiokande, and IMB particle detectors during a 2-hour period near the explosion of supernova SN1987A. In particular, the correlations among the gravitational wave detectors and the Mt. Blanc neutrinos were claimed to have a chance probability of less than  $10^{-6}$ . If this low probability implies that the correlations are a real physical effect, then new physics will be required to explain them. However, one of the statistical tests used to establish these correlations is seriously flawed, and most others were devised *a posteriori* and contain considerable freedom to make choices that affect the probability of finding correlations. By a careful consideration of these free parameters, and by applying similar analysis methods to

---

<sup>1</sup>Note: this chapter is a version of a paper written by C.A. Dickson and Bernard F. Schutz and submitted to Physical Review D in 1992

simulated pseudo-random data sets, we show that the actual frequency with which correlations similar to those in the Mt. Blanc data would occur in random data streams is between 0.1% and 10%. Moreover, if the Mt. Blanc correlations were real, then one would expect them in the other particle detectors. After inspecting the evidence, we also conclude that there are no physically significant correlations of the Mt. Blanc type between the gravitational wave detectors and the Kamiokande and/or IMB particles. This makes it very likely that the Mt. Blanc correlations are due, not to any physical effect, but simply to chance.

## 2.1 Introduction

At about the time of the supernova explosion SN1987A there were, unfortunately, only two gravitational wave detectors in operation (Amaldi *et al.* 1988). These were of the least sensitive type: room temperature bar detectors, one in Maryland and the other in Rome. There were four proton-decay experiments in operation that had the capability to detect particles from the supernova, and three of them — Kamiokande (Hirata *et al.* 1987), IMB (Bionta *et al.* 1987), and Baksan (Alexeyev *et al.* 1987) — registered a coincident burst. Unfortunately, only one gravitational wave detector was recording data at that time (Rome), and that was affected by seismic noise (Amaldi *et al.* 1988). However, at the time of a somewhat earlier “neutrino” burst in the Mont Blanc detector (Aglietta *et al.* 1987a), which probably was not associated with the supernova, both gravitational wave detectors were working satisfactorily.

Since gravitational waves emitted by the supernova and carrying any reasonable amount of energy would be well below the sensitivity limits of these room-temperature bar detectors, it was not expected that the gravitational wave data would show any signals. The first published analyses by the teams involved in the detection and analysis of the data, to whom we shall refer as the Rome-Turin-Maryland Collaboration (RTM)<sup>2</sup>, found: (i) that with a delay of 1.4 s with respect to the 5 neutrino events of the apparent burst, the Rome gravitational wave data were at an appreciably higher level of excitation than average (in particular, there was an unusual excitation of the Rome detector just before the first Mont Blanc neutrino event (Amaldi *et al.* 1987), with a chance probability of 3%); and (ii) there

---

<sup>2</sup>Our use of the abbreviation “RTM” is a convenient shorthand for a large team of authors: not all individual authors contributed to all papers; and in two cases, Aglietta *et al.* (1989); Aglietta *et al.* (1991b), some authors from Moscow have contributed. See the individual citations for the full author lists.

was a modest correlation between the Rome and Maryland gravitational wave detectors in a 7 hour period around the time of the Mont Blanc burst (Amaldi *et al.* 1988) (with chance probability 3.5%). But they reported no unusual coincidences between the two gravitational wave detectors just at the time of the Mont Blanc burst. On this evidence, there would be no reason to suppose that gravitational waves from SN1987A had been detected.

However, in subsequent analyses, RTM searched a larger stretch of data for further events like those reported earlier (Amaldi *et al.* 1987), where a gravitational wave detector is excited a fixed time before a particle is detected. This has led to a series of papers (Amaldi *et al.* 1987; Aglietta *et al.* 1989; Amaldi *et al.* 1989; Aglietta *et al.* 1991b; Aglietta *et al.* 1991a) reporting that time-delayed coincidences have occurred in various stretches of data with apparently high significance (low chance probability). RTM have found numbers of delayed coincidences between the gravitational wave detectors and the Mont Blanc neutrino detector (Amaldi *et al.* 1987; Aglietta *et al.* 1989) and between the gravitational wave detectors and the Kamiokande (Amaldi *et al.* 1989; Aglietta *et al.* 1991b), Baksan (Aglietta *et al.* 1991b), and IMB (Aglietta *et al.* 1991a) particle detectors respectively. RTM assigned chance probabilities to various of these coincidences in the range from  $10^{-2}$  down to  $10^{-6}$ . Our main purpose in this paper is to reassess these claims by RTM.

It seems clear that if the delayed coincidences are due to a real physical effect, then new physics will be required to explain them. Tens of coincident events are claimed to have taken place over a 2-hour period. If they are due to neutrinos and gravitational waves from SN1987A, the energy involved would be huge, many thousands of solar rest masses converted into gravitational wave energy *for each event*<sup>3</sup>. Moreover, given the low efficiency of neutrino detection, potentially thousands of events may have been missed. If they are not gravitational waves and neutrinos, then some new particles and interactions would be required.

One's attitude towards the need for new physics depends on (i) the *significance* of the observed correlations and (ii) one's assessment of the *plausibility* of the new physics required<sup>4</sup>.

---

<sup>3</sup>The claim that gravitational wave detectors actually have a much larger cross-section than we have taken here — e.g., Weber (1984) — is wrong, as has been shown by Thorne (1992) and Grishchuk (1992); and so does not offer a way out of these problems.

<sup>4</sup>It would be interesting to take a more Bayesian approach to the whole analysis (see, e.g., Bulmer 1979), perhaps by assigning some *a priori* probability to the new physics required. Because RTM do not postulate any consistent physical explanation of their correlations, we shall not attempt this.

In practice, the significance of unusual results is usually taken to be the *a priori* probability of obtaining the results under a null hypothesis (normally that the data are completely random). The significance is particularly difficult to assess when, as here, the correlations were unexpected and so were only found *a posteriori*, after examination of at least part of the data set, and even then only by unusual statistical methods. RTM understand this and attempt to take account of the *a posteriori* nature of the effect by using other stretches of data as “control” sets in which to look for chance correlations. Unfortunately, we shall show that their “control” analysis is seriously compromised by the way the control data were chosen and by the statistical dependence of data they treat as independent. In fact, their principal statistical test is so flawed by data dependencies that we believe it is impossible to draw reliable conclusions from it.

We therefore undertake as part of our analysis to provide a more reliable control set by generating large numbers of random data sets on a computer and using RTM’s own methods to analyze them. We directly address the question of how much freedom RTM had to find *a posteriori* effects in their original analysis of the Rome–Maryland–Mont Blanc data, such as by varying the time-delay and the thresholds of the gravitational wave detectors until they found significant coincidences. (RTM do in fact describe doing this.) Our approach cannot, of course, do more than estimate the true chance probability of the correlations, but it is a completely independent analysis, and it gives a radically different answer from the one RTM give.

Regardless of the significance of the Rome–Maryland–Mont Blanc correlations, the acid test of whether they point to a new physical effect is whether similar correlations occurred between the gravitational wave detectors and other particle detectors at the same time. RTM analyzed the data from the Kamiokande (Amaldi *et al.* 1989), Baksan (Aglietta *et al.* 1991b) and IMB (Aglietta *et al.* 1991a) detectors and claimed that they do in fact support the reality of the effect: they find correlations which they claim are very significant. Unfortunately, their analyses are again compromised by their data-selection criteria, by time-keeping problems in two of the detectors involved, and most seriously by the fact that, as we shall show, *RTM do not find significant correlations when they analyze the data in the same way as they analyzed the Mt. Blanc data.*

RTM themselves admit that, using these analysis techniques, there are no significant correlations between gravitational waves and Baksan particles (Aglietta *et al.* 1991b). They

find modest correlations between gravitational wave data and IMB and Kamiokande particle events using the same analysis techniques, but we shall show that their analysis is fatally compromised by various data-selection criteria and by time-keeping problems. They find apparently significant correlations in these three detectors only when they use new methods of analysis not applied to the Mt. Blanc data. It is our conclusion that there is no evidence of correlations of the Mt. Blanc type in the Kamiokande, Baksan or IMB data, and that therefore the RTM correlations fail this crucial predictive test.

The data and analyses of RTM appear in a number of places in the scientific literature, some of them not widely available. We therefore shall try to make this paper as self-contained as possible. We begin in Section 2.2 with a review of the actual observations made by the two gravitational wave detectors and two particle detectors at the time of the supernova. In Section 2.3 we then give a summary of RTM's main analysis techniques. We point out that one of their unusual methods (which we call the net-excitation method) is seriously flawed. In Section 2.4 we present our own analysis of simulations of the Mt. Blanc and gravitational wave data streams, using mainly the other RTM method (the threshold-coincidence method). We find that the frequency distribution in random data sets of the sorts of correlations that RTM find is very much larger than RTM estimate. This allows us to make a detailed reassessment in Section 2.5 of the coincidence claims, including an attempt to correct for the large number of sometimes hidden degrees of freedom that have been used by RTM to optimize the correlations. These include the following:

1. *a posteriori* choices of, or freedom to choose, the time-delay
2. *a posteriori* choices of, or freedom to choose, the gravitational wave threshold
3. choice and variation of the duration of the data set
4. choice and variation of the starting time of the data set
5. statistical dependence of data sets caused by including the original "eyeballed" data set in the larger ones that were subjected to an analysis that was based on inspection of the original set
6. use of nonstandard and seriously flawed statistical tests with poorly-understood statistics, when standard tests could have been used but were not (or were not reported);  
and

7. the failure to apply consistently the Mt. Blanc analysis methods to data from Kamiokande and IMB.

(Some of the details of RTM's analysis are deferred to the appendix, with additional criticism where appropriate.)

The effects of some of these degrees of freedom are fairly easily quantified, while some are not so easily quantified. However, none of them is negligible; and all of them have the effect of making the coincidences more likely to have arisen by chance than RTM have claimed. Our reassessment for the gravitational wave—Mt. Blanc coincidences revises the coincidence probability from  $\sim 10^{-6}$  (RTM's estimate) to  $10^{-3}$ – $10^{-1}$  (our estimate). For gravitational wave—Kamiokande coincidences we revise from  $\sim 10^{-4}$  (RTM's estimate) to the level of chance (our estimate). Finally, for gravitational wave—IMB coincidences we revise from  $\sim 10^{-3}$  (RTM's estimate) to  $\sim 10^{-1}$  (our estimate). We feel that these correlations are therefore much more likely to have arisen by chance than to be a pointer to new physics.

RTM themselves never actually claim that the correlations are due to a real physical effect, and they have not proposed a serious model to explain them. They also remark in places that their probability estimates are only tentative in some respects. Their papers contain full descriptions of the tests that they report, which makes our reassessment possible. However, RTM themselves have not published a more detailed assessment of their significance estimates, and we wish to fill that gap here.

We wish to make clear at this point that it is not the goal of this paper to attempt to give a definitive set of rules of how we believe gravitational wave data *should* be analysed, which is a paper in itself, and which one of the authors will address in his thesis (CAD). However, we could make the following general recommendations:

- that all analyses of a data set, whether or not they give the results expected or desired by the analysers, should be stated;
- that data sets should be carefully examined individually and the results reported before they are combined;
- that the analysis methods used should be standard where possible, and that in any case the statistics of the analysis methods should be well understood or explained, and clear enough to be questioned easily;



- that a clear model should be given and tested (at least, the null hypothesis should always be tested);
- once a new model has been postulated on the basis of a given dataset, any new data should be analysed in the same way as the original data were.

## 2.2 The Gravitational Wave and Neutrino Observations

The observations of SN1987A are well documented (Trimble 1988), so we shall not review all of them here. However, we shall review the observations of the particle and gravitational wave detectors in operation at the time of the supernova.

Note that we have had some difficulty with our nomenclature, not knowing whether events crossing the threshold of a particle detector are neutrinos, some other particle, or random excitations in the detector (a normal background count); and this will vary from one detector to another. To call all the Mt. Blanc events *neutrinos*, for instance, would be presumptuous; and since RTM have still not provided a consistent model for the effect they see, we shall, where appropriate, enclose the word *neutrino* in quotes. For the other three particle detectors, we have generally used the word *particle*; though again, this should not be taken to imply that, in all cases, real particles have been detected, or that the particles are or are not neutrinos.

### 2.2.1 Particle observations

There were four particle detectors in operation during the relevant period: Mont Blanc (variously called UNO or LSD) (Aglietta *et al.* 1987a), Kamiokande (K II) (Hirata *et al.* 1987), IMB (Bionta *et al.* 1987) and Baksan (Alexeyev *et al.* 1987). All four were in operation during the whole of 22-23 February 1987. The optical brightening of the supernova took place between about 2h and 11h UT on 23 February. Neutrinos would have been expected at any time up to 24 hours before this, allowing time for the hydrodynamic shock to reach the star's surface and cause the optical display.

At about 2h 52m 37s UT, Mt. Blanc observed a burst of five “neutrino” events (Aglietta *et al.* 1987a). This burst had a probability about  $2 \times 10^{-3}$  of arising purely from the Poisson background during a period of 24 hours immediately preceding the observation of the optical supernova event (Aglietta *et al.* 1987a). However, this observation cannot easily

be reconciled with those of the other detectors in operation since no significant particle bursts coincident with the Mt. Blanc event were observed in the other detectors. Therefore, the Mt. Blanc burst is usually distrusted (Trimble 1988).

The later burst, however, at about 7h 35m UT was certainly a real flux of neutrinos from the supernova: the other three particle detectors in operation all showed signals above the threshold levels about this time. Kamiokande (Hirata *et al.* 1987) detected 11 neutrinos at 7h 35m 35s UT ( $\pm 60$  s) within a time interval of 13 s, with energies between 7.5 and 36 MeV. IMB (Bionta *et al.* 1987) reported 8 neutrinos at 7h 35m 41s UT during an interval of 6 s, with energies from 20 to 40 MeV. Baksan (Alexeyev *et al.* 1987) detected 5 neutrinos at 7h 36m 11s UT (+2 s, -54 s) during a time of 10 s, above an energy threshold of 12.0 MeV.

Mt. Blanc itself did not register an intrinsically significant burst at this time, although it did record two events at 7h 36m 00.5s UT and 7h 36m 18.9s, discovered in the off-line analysis (Aglietta *et al.* 1987b). This is not particularly worrying: since Mt. Blanc is smaller than KII and IMB, one would only have expected of the order of 1.5 neutrinos.

We have indicated above a very important point for our analysis, namely that two of the particle detectors had serious uncertainties in the offset of the experiment's clock relative to Universal Time: Kamiokande (Hirata *et al.* 1987) had an absolute timing uncertainty,  $\Delta t_K$ , of  $\pm 60$  s; while the absolute uncertainty  $\Delta t_B$  in the Baksan clock (Alexeyev *et al.* 1987) lay in the range  $-54 \text{ s} < \Delta t_B < 2 \text{ s}$ . The absolute timing of the other two detectors was more accurate, with Mt. Blanc (Aglietta *et al.* 1987a) accurate to  $\pm 2$  ms and IMB (Bionta *et al.* 1987) to  $\pm 50$  ms. The relative timing accuracy between particle events in any given detector was extremely good: the only uncertainty is the constant time shift between the detector clocks.

Given the fact that all three events were well above threshold and that the timing uncertainty allows them all to be coincident, there is little doubt that they are supernova neutrinos. However, the timing uncertainty makes it difficult to assess the probabilities of any coincidences between these detectors and the gravitational wave detectors. We shall return to this point in Sections 2.3 and 2.5.

### 2.2.2 Gravitational wave detectors

The Rome and Maryland room temperature bar gravitational wave detectors operated satisfactorily at least from 18h 24m 3s of 21 Feb 1987 to 6h 2m 3s of 23 Feb 1987, a period that

includes the Mt. Blanc burst but excludes the time of the KII-IMB-Baksan events. Soon after 6h on 23 February, the Maryland detector experienced electrical problems; and at 7h 35m UT, the time of the KII-IMB-Baksan coincident burst, there were seismic disturbances in Rome. RTM confine all their analyses to the period before 6h 2m 3s on 23 February, when both gravitational wave detectors were working.

The Rome antenna has a mass of 2300 kg and a resonant frequency of 858 Hz. The Maryland antenna has a mass of 3100 kg and a resonant frequency of 1660 Hz.

The data sampling rate of the Rome detector was 1 Hz, while that of the Maryland detector was 10 Hz. In order to compare the two data sets, RTM averaged the Maryland data over 1 s intervals. This is 3 times longer than the optimum averaging time for this antenna, so the resulting data set has poorer than optimum signal-to-noise ratio by a factor of  $\sqrt{3}$ .

Before 6h, the gravitational wave detectors seem well-behaved. Events in both detectors followed fairly well an exponential (thermal) distribution in energy, although both detectors had some extra events at higher energies (Aglietta *et al.* 1989). RTM should, perhaps, have performed a more thorough investigation of the data from the individual detectors. The mean noise temperatures were approximately 28.6 K (Rome) and 29.8 K (Maryland).

The Maryland clock maintained an accuracy of  $\pm 0.1$  s during this period. The Rome clock did have an error, but careful study of its behavior after the end of the observation period led RTM to apply a correction of  $(-0.7 \pm 0.1)$  s to obtain the true time.

## 2.3 Summary of the Main RTM Analysis Methods

Here we review the main methods of the RTM coincidence analysis.

### 2.3.1 The main RTM analysis methods

#### The RTM “net excitation” method

The first method is to sum the values of the combined gravitational wave streams at all “coincidence times”, namely the arrival times of the “neutrinos” minus a fixed chosen time-delay. While this method is unusual, it is not necessarily implausible; however, its statistics are obscure. RTM assess the statistics by examining the behavior of their data set under simple modifications of the method, such as changing the time-delay. We shall see that there

are serious difficulties with the manner in which they do this.

Calling the energy excitations of the Rome and Maryland antennae  $E_R(t)$  and  $E_M(t)$ , respectively, the principal statistic used by RTM in their first analysis method is what we shall call the “net excitation” of the gravitational wave detectors over this period:

$$C_*(\phi) = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} [E_R(t_i + \phi) * E_M(t_i + \phi)], \quad (2.1)$$

where  $\phi$  is a chosen offset time,  $t_i$  is the arrival time of the  $i$ -th “neutrino”,  $N_\nu$  is the total number of “neutrinos”, and “\*” indicates either “+” or “ $\times$ ”, depending on whether one is using the sum or product of the gravitational wave signals. When the offset  $\phi$  is negative we shall refer to it as a *time-delay* (of the “neutrinos” relative to the gravitational waves), and an *advance* when positive. The values of  $t_i + \phi$  are rounded to the nearest gravitational wave sampling time for the evaluation of  $E$ .

When adding the signals (“\*” = “+”), one has to decide how to weight the two detectors. (This does not apply to the product algorithm, but most of RTM’s analyses, including *all* their most improbable correlations (Aglietta *et al.* 1989; Amaldi *et al.* 1989; Aglietta *et al.* 1991b; Aglietta *et al.* 1991a.) use the sum algorithm only.) The decision of RTM (Aglietta *et al.* 1989) is to normalize them by the mass of the detector, i.e., to divide the energy of the Maryland antenna by the ratio 3100/2300 of the mass of the Maryland detector to that of the Rome detector. This is somewhat arbitrary, since it takes no account of the large difference in the resonant frequencies of the two antennas, which implies that they respond to completely different parts of the spectrum of any gravitational wave event. Note that RTM also do *not* make any correction for the different orientations of the detectors.

RTM assess the significance of any result by comparing  $C_*(\phi)$  with some “background” values of the same quantity, as determined by using different time delays in the two gravitational wave streams:

$$C_*(\delta_1, \delta_2) = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} [E_R(t_i + \delta_1) * E_M(t_i + \delta_2)], \quad (2.2)$$

where  $\delta_1$  and  $\delta_2$  are separate time delays. We shall see in a moment that this definition of a comparison background fatally flaws this method.

By changing the time delays, RTM calculate a large number  $N_b$  of these background values, between  $N_b = 10^3$  and  $N_b = 10^6$  in various investigations. They then assign a

ranking order to the various time-delays  $\phi$  by defining

$$n(\phi) = \text{count}_{\delta_1, \delta_2} [C_*(\delta_1, \delta_2) \geq C_*(\phi)], \quad (2.3)$$

where “count<sub>range</sub>[ < condition > ]” means that one counts the number of times the condition is true for variables in the given range. In this case, the smaller is the value of  $n(\phi)$ , the more significant is the correlation for that time-delay. Since the range of  $\delta_1$  and  $\delta_2$  always includes  $\phi$ , the minimum value of  $n(\phi)$  is 1. The maximum is  $N_b$ .

On the assumption that the background values are all independent, the probability of the correlation at a given delay is then taken by RTM to be

$$p(\phi) = n(\phi)/N_b. \quad (2.4)$$

If the background values  $C_*(\delta_1, \delta_2)$  were all independent and had the same distribution as the values of  $C_*(\phi)$ , and if  $n(\phi) \gg 1$ , then this would not be an unreasonable way of estimating the probability. Unfortunately, none of these three conditions holds in the RTM analysis. We shall examine the independence of the background values in a moment. (We discuss the effect of small-number statistics [ $n(\phi) \sim 1$ ] in Appendix 2.7.1, and we return to the question of the distributions of  $C_*(\phi)$  and  $C_*(\delta_1, \delta_2)$  in Appendix 2.7.2.)

Notice that this method uses only the ranking order of the values of the correlations, and does not attempt to use a frequency distribution in uniform steps of  $C_*$ , which would be more conventional. This means that two values of  $\phi$  may give values of  $C_*$  that are very close, but they could be far apart in  $n(\phi)$ .

### Criticism of the net excitation method

The biggest problem with the net excitation method is that the background values are not all independent. This is easy to see if we count the number of data points from which the background values are derived. RTM say that they always take values of  $\delta_i$  such that the background value is taken from the same period as the signal,  $C(\phi)$  (Aglietta *et al.* 1989). This is to avoid problems due to possible nonstationarity of the noise. Now, in a 2-hour stretch of data, where RTM find their strongest correlations (Aglietta *et al.* 1989), there are 7200 1-second samples from each gravitational wave detector. On the null hypothesis (no genuine correlation), there are thus about  $1.4 \times 10^4$  independent random numbers in the original data. These numbers are combined in various ways using Eq. 2.2 to form up to  $10^6$

background values. There must, therefore, be hidden correlations among the background values, at least when  $N_b$  exceeds about  $10^4$ . It would not be easy to characterize these correlations, but it would be most unwise to assume, as RTM do, that there are none of significance for this method. Any estimate of probability from this method below a few times  $10^{-4}$  cannot, therefore, be reliable.

Indeed, we shall see that the results of this test, as reported by RTM, show great variations in the apparent probabilities for time-delays separated by as little as 0.1 s, well below the physical resolution of the gravitational wave experiments. This may well be due to the untrustworthiness of Eq. 2.4 for the smallest apparent probabilities.

### Threshold coincidence method

The second RTM method is similar to the threshold-crossing gravitational wave–neutrino method we suggested at the beginning of this section, only it is applied to the combined gravitational wave data stream rather than to each one separately. RTM set a threshold on the combined data stream

$$E_*(t) = E_R(t) * E_M(t) \quad (2.5)$$

(where again “\*” is “+” or “×”), and identify gravitational wave “events” as those which cross the threshold. (These are not of course necessarily real gravitational waves: they may be just thermal noise excitations.) A coincidence occurs for a time-offset  $\phi$  with a “neutrino” that arrived at time  $t$  if the event occurs at the nearest gravitational wave sampling time to  $t + \phi$ . The statistics of this method are much more straightforward, at least for a fixed threshold.

For a data set lasting  $N_t$  sampling intervals (of one second), containing  $N_\nu$  “neutrinos” and  $N_{gw}$  gravitational wave events randomly (uniformly) distributed, the expected number of coincidences is

$$\bar{n} = \frac{N_\nu N_{gw}}{N_t}. \quad (2.6)$$

Given that arrival times are uniformly distributed, the probability of obtaining  $n$  or more coincidences, given the mean  $\bar{n}$ , is

$$p_{\bar{n}}(n) = \sum_{r=n}^{\infty} \frac{\bar{n}^r e^{-\bar{n}}}{r!} = 1 - \sum_{r=0}^{n-1} \frac{\bar{n}^r e^{-\bar{n}}}{r!}. \quad (2.7)$$

This equation holds provided  $|\phi|$  is much less than the expected interval between coincidences; if  $|\phi|$  is too large, end effects will reduce the coincidence probability.

## 2.4 Monte Carlo simulations

### 2.4.1 Computer model

The objective of our Monte Carlo computer simulation was to assess the realistic probability that the correlations found by RTM would arise by chance in completely random data sets. With computer-generated data we can experiment with changing thresholds, time delays, and even methods of analysis to see what effect these have on apparent correlations. We have generated large numbers of pseudo-random data, analyzed them using the RTM threshold-coincidence method, computed the apparent probability of the strongest correlations by RTM's net-excitation method, and then compared this with the actual relative frequency of occurrence of such correlations among the pseudo-random data sets. We principally simulate the analysis of the Mt. Blanc data, although our results also illuminate the treatment of the Kamiokande and IMB data.

### 2.4.2 Properties of the pseudo-random data

In each Monte Carlo run, two sets of artificial gravitational wave data were generated, one corresponding to the energy excitation of the Maryland detector and the other to that of the Rome detector. Each artificial data set consisted of 7200 samples, equivalent to a 2-hour data stream sampled at 1 s intervals. The samples were drawn from distributions which were exponential in the temperature of the excitation, the Rome simulated data with mean 28.6 K and the Maryland with mean 22.1 K (its effective temperature after normalizing its mass to that of the Rome antenna and averaging over 1 s intervals for comparison with the 1 Hz Rome data (Aglietta *et al.* 1989)).

For the “neutrinos”, we assumed an exponential distribution of the time delays between one neutrino and the next, using the observed mean arrival interval in the Mt. Blanc data (Aglietta *et al.* 1989). (This is the distribution one expects, of course, if the neutrinos arrive according to the standard Poissonian “shot noise” model.)

To generate the random numbers we used the *Numerical Recipes* (Press *et al.* 1988) uniform random number algorithm *RAN1*. The cycle length of this random number generator is said to be infinite for all practical purposes (Press *et al.* 1988). We demonstrate its distribution by generating and binning the first  $4 \times 10^5$  numbers in Fig. 2.1.

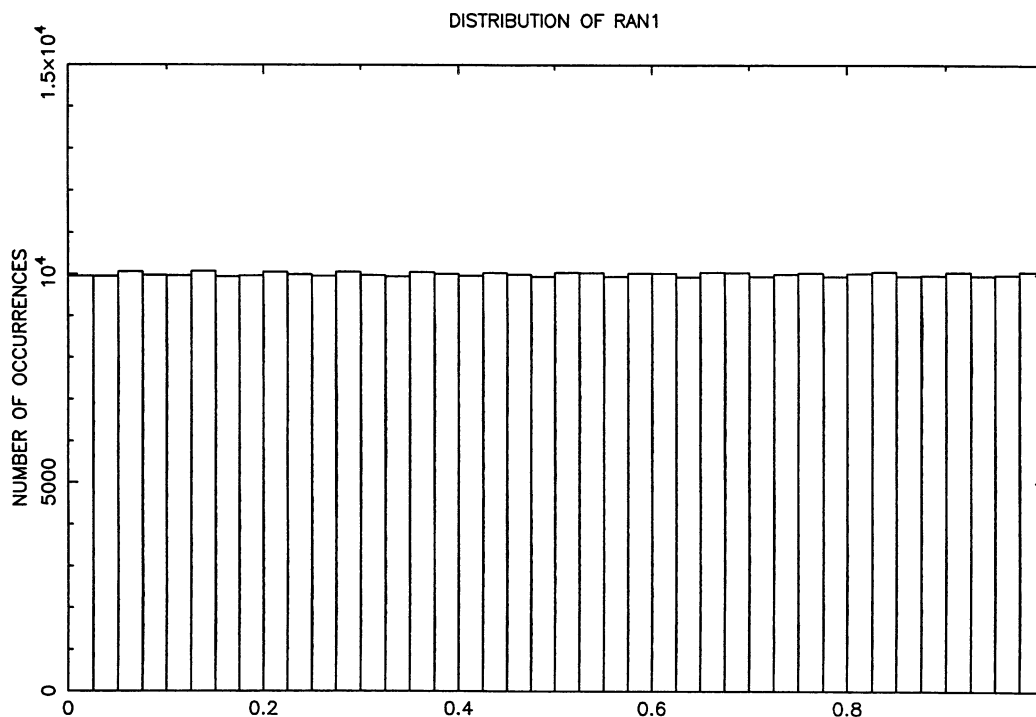


Figure 2.1: Distribution of the pseudo-random number generator

### 2.4.3 Method of analysis of the pseudo-random data

#### Adoption of the threshold-coincidence method

To analyze the simulated data one needs to choose from the large variety of statistical tests which RTM employ. Since the main thrust of this paper is to examine the Mt. Blanc-gravitational wave coincidences, we shall perform an RTM-style analysis of two artificial gravitational wave data streams and one artificial neutrino stream.

We perform an RTM-style threshold-coincidence analysis on each random set. This allows us to assess the influence of the freedom to choose the best threshold on the chance of finding a strong correlation. In view of the dubious value of the net-excitation measure of correlations, it would be inappropriate to subject each random set to such an analysis. Indeed, the computer time that would be required for such an analysis would be huge, since millions of random numbers would be required for the analysis of each data set. (Once a given set is generated, one needs to generate from it all the background values.) Instead, only for any data sets in which we found significant threshold-correlations do we also perform a net-excitation analysis. We will see that this still sheds considerable light on the question



of how unusual are the correlations whose claimed (apparent) probability is  $10^{-6}$ .

For each Monte Carlo data set, we have two choices to make, the threshold and the time-delay. We shall discuss each of these choices in turn.

### Selection of a threshold

In choosing the threshold  $T$ , we are guided by what RTM say about their choice (Aglietta *et al.* 1989). They select  $T = 150$  K for the summation statistic in the net-excitation method because it gives the best correlation. They indicate that they searched thresholds from 100 K to 200 K in steps of 10 K. In our simulations, therefore, we search through the same set of thresholds. This is a minimal set: we can be confident they searched all of these. If in fact they searched a larger number than they displayed in Fig. 16 of Aglietta *et al.* (1989), then the “true” probability of a correlation would be larger.

### Choosing a time delay

Although it is clear that RTM searched some range of time delays before settling on their preferred one of 1.1 s, the central problem for us is to decide how wide that range should be when we analyze our simulated data. Note that, despite our reservations about the wisdom of varying time-delays in steps of only 0.1 s when the gravitational wave data have a time-resolution of 1 s, we must follow RTM in this if we are to simulate their methods faithfully.

In analysing the Mt. Blanc data, RTM changed their “best” delay from 1.4 s (Amaldi *et al.* 1987) to 1.2 s (Aglietta *et al.* 1989) and then to 1.1 s (Aglietta *et al.* 1989), depending which was the optimum delay for the data under consideration and the analysis method in question, so some *a posteriori* adjustments were made. RTM thus indicated their willingness to optimise the time delay, within a not-well-defined range, on receipt of more data and the use of other analysis methods. (In the case of the net excitation in Aglietta *et al.* (1989), this optimisation changes the “probability” from  $10^{-4}$ , at delay 1.4 s, to  $10^{-6}$ , at 1.1 s: a large change in “probability” for an apparently insignificant change in delay.) However, RTM never went far from their first value of 1.4 s, which they found by inspection from the raw Rome and Mt. Blanc data at about the time of the  $5\nu$  burst. RTM also tell us that they would never have adjusted the delay by more than about 0.5 s–1 s (Pizzella & Pallottino 1991), although of course this comment was made after publishing the results.

The crucial question for us is the following. Given that the initial eyeballing of the data had provided the motivation to search time-delays, if it then happened that, after receipt of the Maryland data and a full analysis of both gravitational wave datasets, RTM had discovered a much stronger correlation at a very different delay, would they have ignored it? Would they have been bound by their original choice of  $1.4 \pm (\text{say}) 0.5\text{s}$  when the phenomenon, by hypothesis, occurs over a period of 2 hours in both data sets, and when their original choice was made simply by crude eyeball inspection of 50 s of one of the data sets? We believe that, had a much better delay been found, RTM *should* have rejected their original choice completely.

Moreover, RTM *did* in fact search a wide range of time-delays after receiving the Maryland data. Using the net excitation method, they looked at delays from  $-3.2\text{ s}$  to  $+0.8\text{ s}$ , which was not necessary for the calculation of the strength of the correlation at  $-1.4\text{ s}$ , which they had postulated. It was this search that led to their later adoption of a delay of  $1.1\text{ s}$ . Also, using the threshold coincidence method, RTM searched from  $-50\text{ s}$  to  $+50\text{ s}$ , for a fixed threshold, and found no correlations stronger than those around  $1.2\text{ s}$ . This is not surprising since the threshold was optimised for the chosen delay of  $1.2\text{ s}$ . If, in either of these searches, they had found any correlations which were stronger yet, RTM would surely have been obliged to take them seriously.

We conclude, therefore, that we should search our simulated data sets over a wide range of delays. This view is reinforced by an examination of RTM's initial selection of a  $1.4\text{ s}$  delay.

**On the initial selection of the  $1.4\text{ s}$  time-delay.** RTM initially inspected a small stretch of data containing the 5 Mt. Blanc “neutrinos” (see our Fig. 2.2; only the “neutrinos” and the Rome data were used), and they chose a delay for which the gravitational waves are “in most cases appreciably higher than the average background” (Amaldi *et al.* 1987). Since this criterion is just an “eyeball” implementation of their own net excitation method (Eq. 2.2) adapted for one detector instead of two, we shall now use this method to attempt to quantify the effect of their inspection process.

The first RTM time-delay estimate involved only the Rome data, so in Fig. 2.3 we plot the statistic

$$C(\phi) = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} E_R(t_i + \phi), \quad (2.8)$$

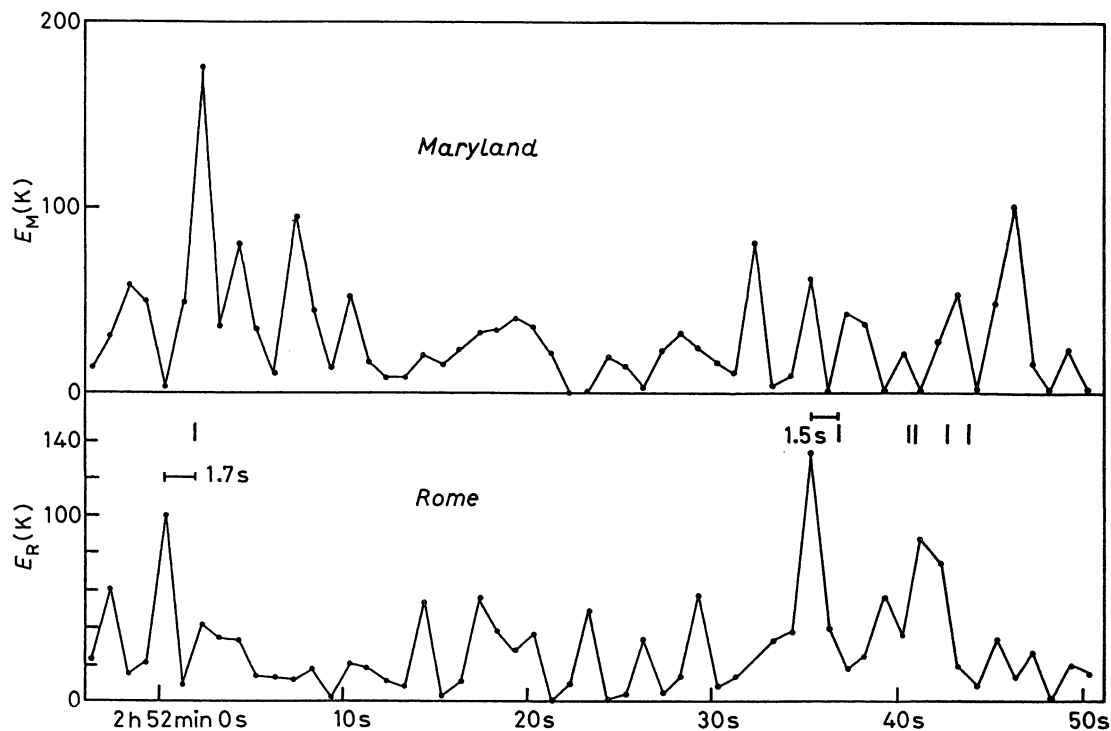


Figure 2.2: First indications of the Mt. Blanc correlations. RTM originally had only the Mt. Blanc “neutrinos” and the Rome data from which to select a delay of 1.4 s. The Maryland data were obtained later, and appear in the analyses in Amaldi *et al.* 1988. (Reproduced from Aglietta *et al.* 1989 with permission.)

which is the single-detector version of  $C_*$  of Eq. 2.1. Our figure contains two plots: (a) uses all six “neutrinos” that are shown in Fig. 2.2; (b) uses only the five “neutrinos” of the Mt. Blanc burst. In both cases, the best time-delays are between 1.3 s and 1.8 s, but there is no preference among them. This agrees with the RTM choice. But other delays offer hope of some effect: near 5.5 s and 7.5 s there are peaks above 50 K. Note that each peak is about 1 s wide, which agrees with the time-resolution of the gravitational wave data. When RTM broadened the analysis from the Rome–Mt. Blanc to the Rome–Maryland–Mt. Blanc data, they changed the delay from 1.4 s to 1.2 s after a similar eyeball inspection of a short stretch of the data (Amaldi *et al.* 1987; Aglietta *et al.* 1989). Accordingly, we next look at the effect on the delay when we include the Maryland data. Thus, we next look at the full net-excitation statistic  $C_+$  (Eq. 2.1) applied to all the data of Fig. 2.2. Our results are shown in Fig. 2.4. Here, the picture is very different: there is little to choose between time delays

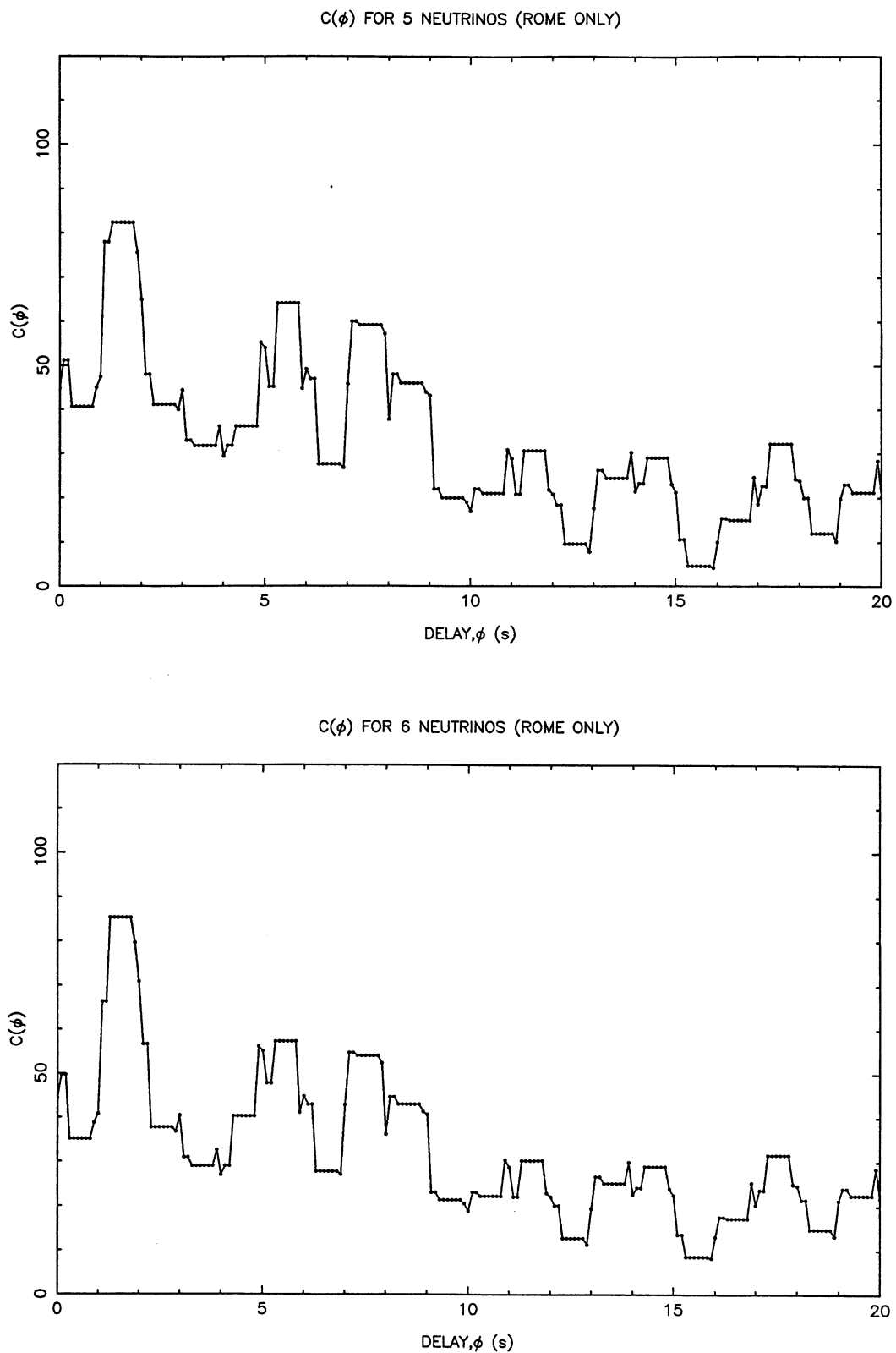


Figure 2.3: Searching for good time delays using the net-excitation method applied to the data set of Fig. 2.2, using Rome data only. (a) contains all six “neutrinos” seen in Fig. 2.2, while (b) omits the isolated “neutrino” event near 2h 52m 2s.

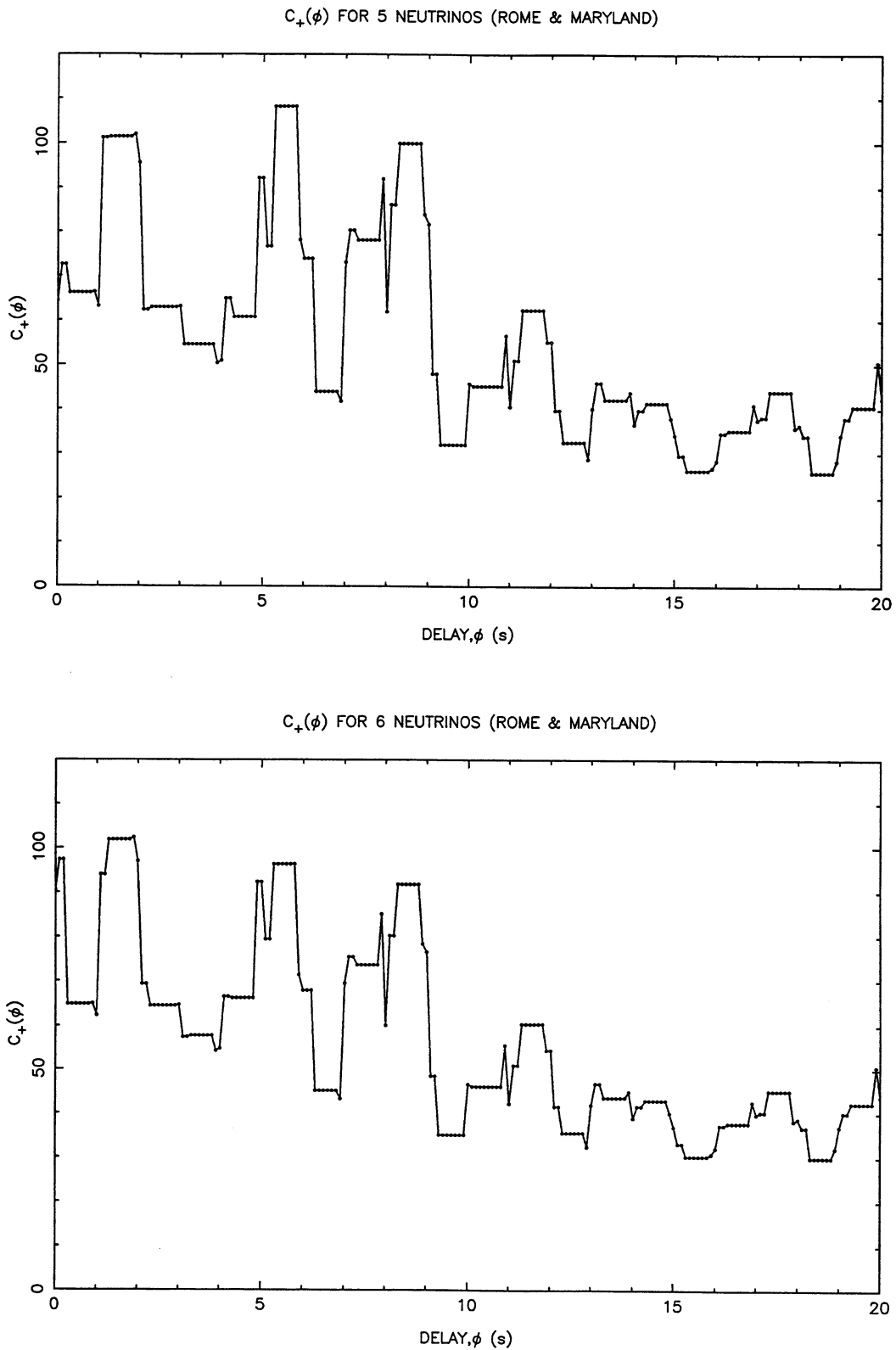


Figure 2.4: Searching for good time delays using the net-excitation method applied to the data set of Fig. 2.2, using both Rome and Maryland data. (a) contains all six “neutrinos” seen in Fig. 2.2, while (b) omits the isolated “neutrino” event near 2h 52m 2s.

near 1.5 s and those near 5.5 s and 8.5 s. In fact, if one uses only the 5 “burst neutrinos”, the *best* time-delay is 5.5 s.

One could argue, therefore, that based on RTM’s own selection criterion, they could have changed the delay time from 1.4 s to 5.5 s on receipt of the Maryland data. In fact, to us it seems natural to try to match the “double neutrino” event (arrival times 40.6 s and 41.0 s) with the highest Rome peak at 35.3 s, leading to a delay of about 5.5 s, which as we have seen is as good as or better than the delay they originally chose.

This is not, of course, to argue in favor of the reality of correlations at other delays. Our point here is to show that the range of time-delays that were open to RTM was considerable. Had, by accident of the noise in the gravitational wave detectors, the time-delay at 5.5 s proved a bit more significant, RTM would presumably have had no problem justifying its adoption. The physical model that they offered as a possible justification for the 1.4 s delay — that a small neutrino mass delays them relative to the gravitational waves — is untenable on other grounds (see our Section 2.1), and in any case it could surely have been stretched to justify a 5.5 s delay. Other ad hoc models, perhaps invoking unknown particles that excite the gravitational wave detectors, could easily have been devised to justify either advances or delays of small or moderate size.

We believe, therefore, not only that much larger values of  $|\phi|$  could have been defended, but, indeed, that they *should* have been thoroughly examined by RTM once a time-delay model was adopted for analysis. As we have seen, RTM did indeed perform such an examination.

**Our choice of delay.** Consequently, we must regard the delay between gravitational waves and “neutrinos” as a free parameter like the threshold, and we choose the most favorable delay (within a pre-determined range) for a given set of random data.

We fix the range of available time-delays by staying with the RTM model (Amaldi *et al.* 1987) of ascribing the delay to the effect of a neutrino mass,  $m_\nu$ . The time-delay between a gravitational wave traveling at the speed of light and such a neutrino with energy  $E_\nu$  after traveling a distance  $d$  is

$$\delta t = \left( \frac{m_\nu c^2}{E_\nu} \right)^2 \frac{d}{2c}. \quad (2.9)$$

We need only fix an upper bound on the allowed mass and adopt a value for the typical energy of the neutrinos. By changing RTM’s value of 10 eV for the maximum neutrino mass

(Amaldi *et al.* 1987) to a still reasonable 20 eV, and by relaxing the RTM energy estimate of 10 MeV to the actual measured average energy of the five Mt. Blanc burst events (8.4 MeV), we broaden RTM's allowed range of (0, 2.7) s to (0, 15.3) s. Hence we have run our main Monte Carlo experiment with the choice of delays

$$0.0 \text{ s} \leq \delta t \leq 14.9 \text{ s}, \quad (2.10)$$

in steps of 0.1 s.

For this parameter we feel we may have been conservative, i.e., that we could have defended wider ranges and hence obtained even larger corrected probabilities for the correlations. One could argue that negative delays (neutrinos preceding gravitational wave “events”) should have been considered, since the new physics required to explain any correlations might well involve a new elementary particle that excites the gravitational wave antennas, and this might have traveled more slowly than the neutrinos. By the same argument, the time delay between neutrinos and the new particle could have been very much greater than the limits from the mass of the neutrino, since the new particle's mass could be very much larger. Without an *a priori* model for the physics of these correlations, it is hard to argue for any restriction on the time delay. Instead, a more practical reason for our accepting the relatively narrow range of 15 s is that RTM would probably not have looked for time delays at all had not the peaks in the gravitational wave stream been fairly near the “neutrinos” in Fig. 2.2.

We shall argue, shortly, that if one adopts a different range, the probability just scales in proportion. For example:

1. (conservative scenario) if one feels that delays in the range (−60 s, +60 s) are suitable, and that this range could reasonably have been searched, then the “true” probability will be larger by about a factor of 8 than the one we derive in Eq. 2.12 below;
2. (RTM scenario) if one feels that RTM's original eyeball estimate was binding,  $\pm 0.5$ s, and that during subsequent analysis no other delay could have been considered, then the “true” probability will be smaller by a factor of about 15 than the one we derive.

This illustrates how hard it is to estimate realistic probabilities when data have been analyzed by *a posteriori* criteria.

Note that in Section 2.5.1 we attempt to calculate the *a priori* probability of the correlations in the 2 hour data set, in a way which is independent of one's guess as to the available

choice of time delay. We do this by removing those 50 s of data which RTM inspected to choose their delay of 1.4 s, and testing the predictive power of this delay on the rest of the two hours of data. We find the results are similar to those in our simulations that use a range of delays of 15 s.

### Our algorithm

Having decided on the ranges of our free parameters, we proceeded as follows. Each simulated data set consisted of two gravitational wave streams and one “neutrino” stream generated as described in Section 2.4.2. For each threshold, we searched through the whole range of time-delays to find which one gave the best correlation as measured by the threshold-coincidence analysis method, and then we calculated the *apparent* probability of this correlation using Eq. 2.7. We performed the same analysis for each allowed threshold, and selected from all the one which gave the smallest apparent probability. We repeated this for each Monte Carlo data set (150 sets in our first run,  $10^4$  in our second) to see how often apparent probabilities smaller than any particular value occur. This allows us to correct the apparent probabilities for RTM’s freedom to choose thresholds and time-delays, a freedom they did not systematically quantify. We assume that the *relative frequency* of any apparent probability in our simulation is the *true* probability that that sort of correlation will arise by chance in a given random set.

### 2.4.4 Results

We performed two simulation runs, the first using 150 data sets and the second with  $10^4$ . We made minor changes between the two, primarily in the range of time-delays we accepted. Because one of the difficulties of understanding the significance of any statistical analysis is knowing what analyses have been performed and *not* reported (Section 2.5.1 below), we report *both* of our analyses here separately. We have not performed any others.

#### First simulation run

In our first run, we permitted the delay to vary from  $-60$  s to  $+60$  s in steps of 0.1 s. Although this range is larger than we have argued for, it is clear that, since each data point in the simulated time-series is independent, the coincidences found for different delays will be uncorrelated if the delays differ by more than 1 s. Therefore the probability of obtaining



a given number of coincidences will simply scale linearly with the number of choices of delay. Searching 150 data sets over a range of 120 s is equivalent to searching 1200 data sets over a range of 15 s, which is the range we adopted for our second run. The first run therefore contains 12% as many independent trials as the second one. We regard one trial that uses a 15 s range of time delays and the range of threshold values described earlier as roughly equivalent to one RTM experiment.

We would therefore expect to find only correlations that have *true* probabilities of the order of  $10^{-3}$  in our first simulation. In fact, we found one data set that had correlations that had an *apparent* “probability” that was even smaller than that of the RTM correlations!

In each of the 150 random data sets we summed the two gravitational wave streams and searched above the selected threshold for coincidences with neutrinos at the appropriate delay. The least probable correlation occurred in data set 55: at threshold 110 K and at delay 28.0 s, we found 22 gravitational wave–neutrino coincidences. In Fig. 2.5 we present these results in the same way as is done in Fig. 14 of Aglietta *et al.* (1989).

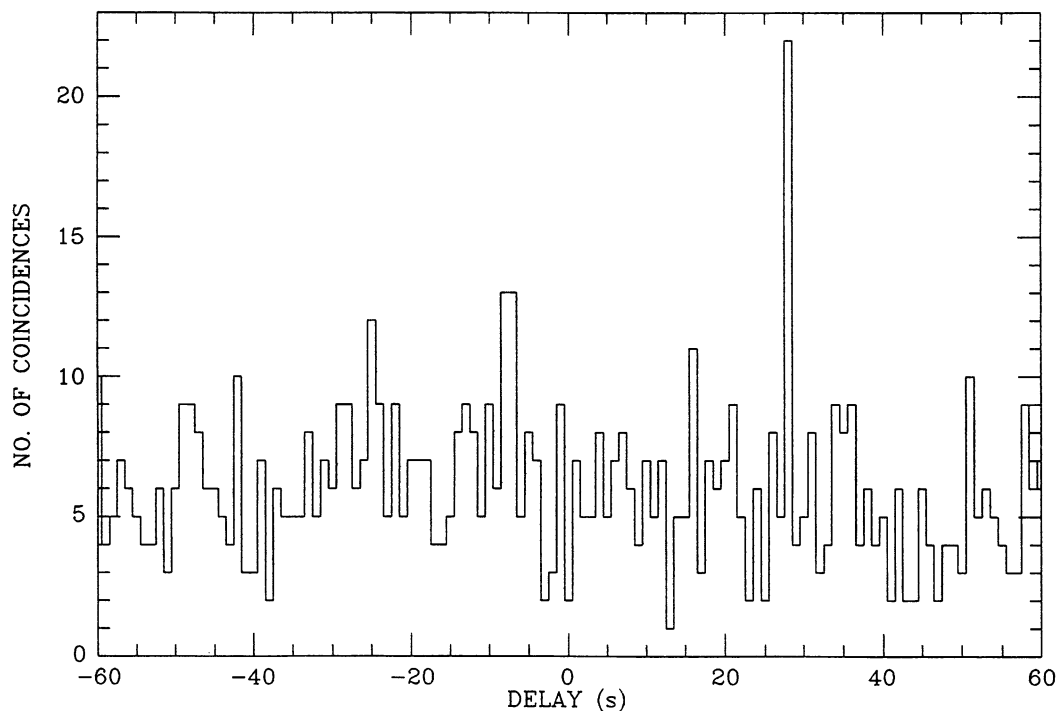


Figure 2.5: Histogram of number of threshold-coincidences against delay time for data set 55 of the first simulation run.

For this data set, there were 86 simulated “neutrinos” within the two hours, and at threshold  $T = 110\text{K}$  there were 512 “gravitational wave events”, giving an expected number of coincidences  $\bar{n} = 6.116$ , by Eq. 2.6. The Poisson probability of obtaining 22 coincidences here is (from Eq. 2.7)

$$p_{\text{lowest}} = p_{\bar{n}=6.116}(22) = 5.3 \times 10^{-7}.$$

This is a *more significant* peak than that found by RTM, using RTM’s method of calculating the probability, although we found it in the equivalent of only 1200 experiments.

We then submitted this data set to a net-excitation analysis, using the summation method and using the “best” time-delay of 28.0 s. A plot of our results in the style of Fig. 11 of Aglietta *et al.* (1989) appears in our Fig. 2.6(a). In Fig. 2.6(b), we reproduce the original RTM figure itself. There is a remarkable similarity between the two. The actual value obtained for  $C_+(28.0)$  was 72.2 K, easily larger than any of the  $10^6$  background values with which it was compared to generate Fig. 2.6(a). We are confident that we could have made the trough in this figure even lower, had we generated more comparison values. We conclude that in roughly 1200 experiments, we have found correlations as strong as than those RTM found in the real data. Note that this was the *first* time we had performed a net excitation analysis, and the only time for these datasets. It is conceivable that there were other datasets in this experiment with net-excitation correlations this strong, and that the threshold coincidence method is an inefficient way of finding them.

However, it is not possible to draw reliable conclusions on the basis of one unusual data set, so we returned to the computer and did a longer simulation run.

### Second simulation run

At the outset of this run we decided that the narrower range of time delays of 0.0 to 14.9 s would be more appropriate for simulating the RTM procedure. We performed  $10^4$  simulations in order to improve our statistics. We still found only one data set which was less probable than the real data, using the threshold coincidence method, but we found several with only slightly larger probability. These have enabled us to form a reliable estimate of the frequency of occurrence of these low-apparent-probability data sets.

**The most improbable simulated data set.** The most improbable data set in our second run was number 327, which had a peak of 9 coincidences at delay 7.6 s at a threshold

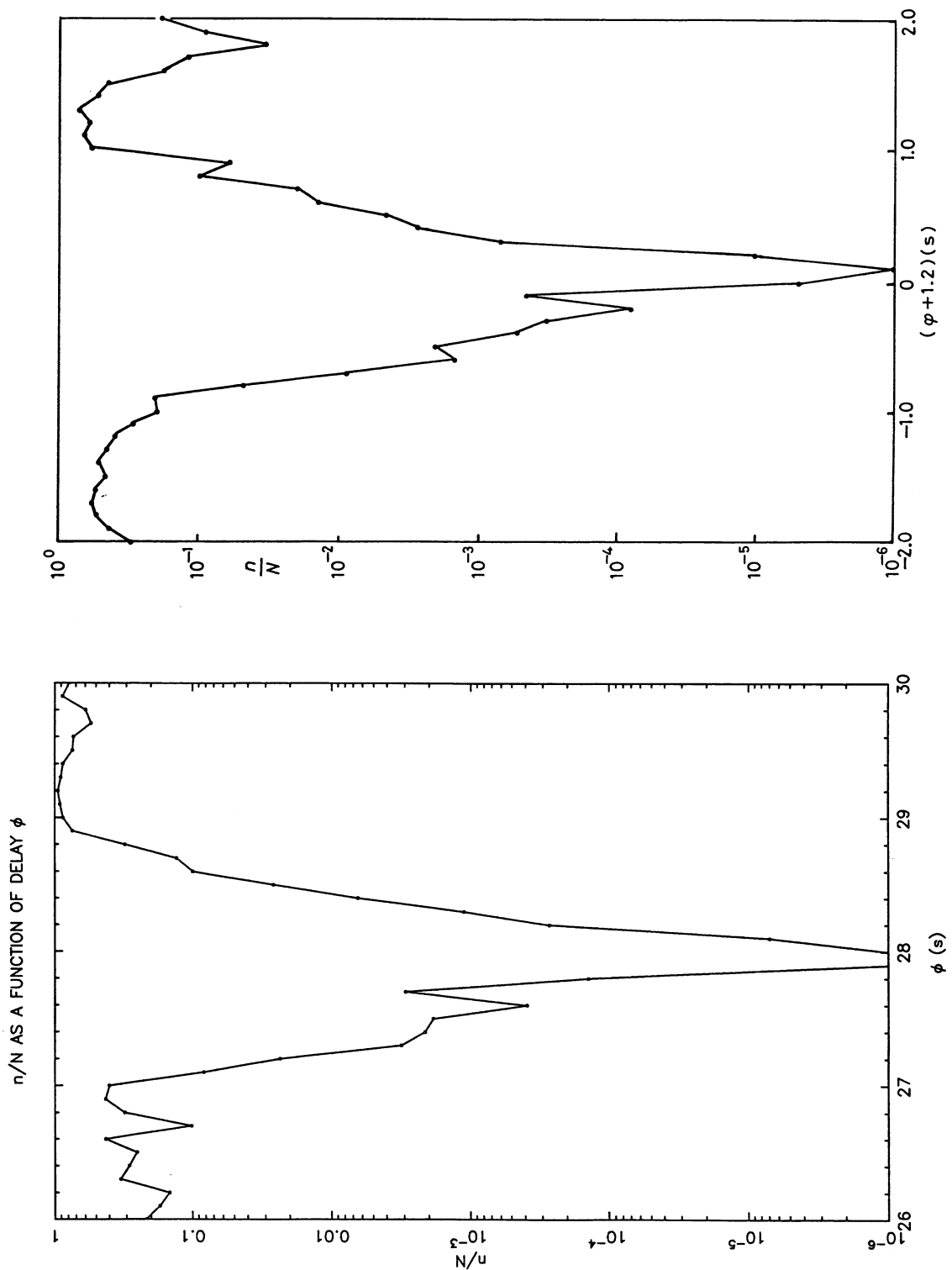


Figure 2.6: Result of the net-excitation analysis of simulation set 55 (a) compared to the RTM analysis of the real “neutrino” and gravitational wave data (b). ((b) reproduced from Ref. Aglietta *et al.* 1989 with permission.)

170 K. The histogram of coincidences against time delays is in Fig. 2.7(a), plotted with the corresponding one for the RTM data (b). There were only 77 neutrinos in the two hours of simulated data, and 83 gravitational wave events above this threshold. The number of expected coincidences is  $\bar{n} = 0.888$  [Eq. 2.6]. From Eq. 2.7, the peak of 9 has a probability of  $4.5 \times 10^{-7}$ , less than that of RTM's correlation.

When we applied the net-excitation analysis to this data set, the result was quite different from that for our earlier data set: the dip in Fig. 2.8 is by no means as dramatic as it was for the RTM data, or for our own Fig. 2.6(a). Although there are an unusual number of coincidences in this data set, the average excitation of the gravitational wave detectors was not extraordinarily high at the (delayed) time of neutrino arrivals. This illustrates simply the fact that the two analysis methods measure different, albeit related, properties of a data set, and so simple probability estimates based on one or another of these statistics will not necessarily agree.

**The relative frequency of occurrence of such correlations.** Given the pseudo-random neutrino and gravitational wave data sets, each threshold  $T$  on the gravitational wave data stream determines an expected number of coincidences  $\bar{n}(T)$ . Choosing a delay  $\phi$  then fixes the actual number of coincidences  $n(T, \phi)$ . We seek the lowest apparent Poisson probability over all thresholds and delays, which we call  $q$ :

$$q = \min_{\text{thresholds } T} \left\{ \min_{\text{time-delays } \phi} p_{\bar{n}(T)}[n(T, \phi)] \right\}, \quad (2.11)$$

where  $p_{\bar{n}}(n)$  is given by Eq. 2.7. The frequency distribution of values of  $q$  in the  $10^4$  data sets gives us our realistic probability distribution. One would expect this to be proportional to  $q$ , if the RTM raw probabilities were realistic, so that smallest values of  $q$  occurred the least frequently. As Fig. 2.9 shows, the actual distribution of  $q$  is just the opposite: the freedom to adjust parameters makes small values of  $q$  very much more probable than large ones.

The analytic form of this distribution is not known, but Fig. 2.10 shows that for most of the range of  $q$  the curve is fairly close to being exponential.

Our interest is in the smallest values of  $q$ , whose histogram is plotted in Fig. 2.11. Within the statistical fluctuations, the distribution is fairly flat, which is what we would expect if the behavior as  $q \rightarrow 0$  is a regular extrapolation to zero of the low- $q$  trend in Fig. 2.10, and does not become singular as  $q \rightarrow 0$ .

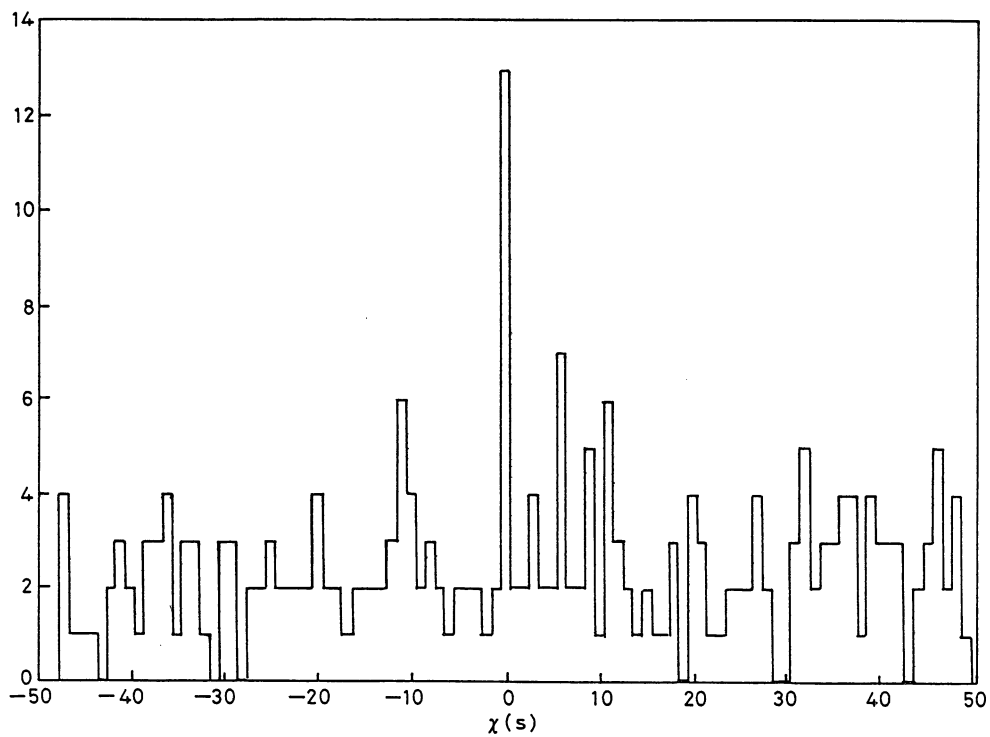
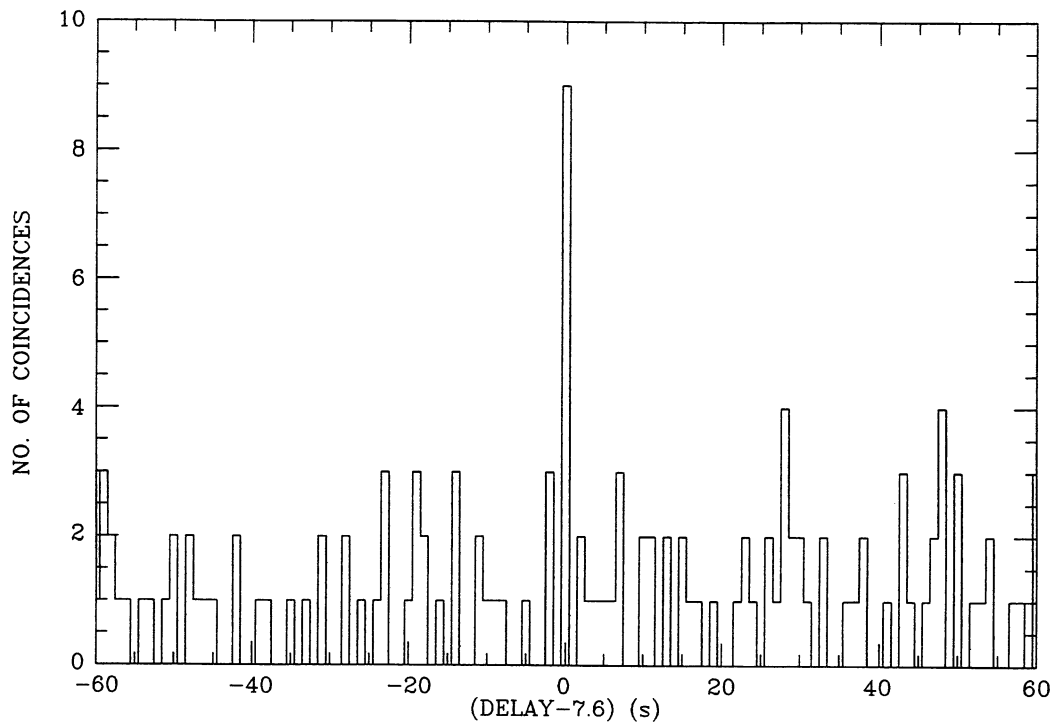


Figure 2.7: Delay histogram for data set 327 of the second simulation (a) compared to the RTM histogram of the real data (b). ((b) reproduced from Aglietta *et al.* 1989 with permission.)

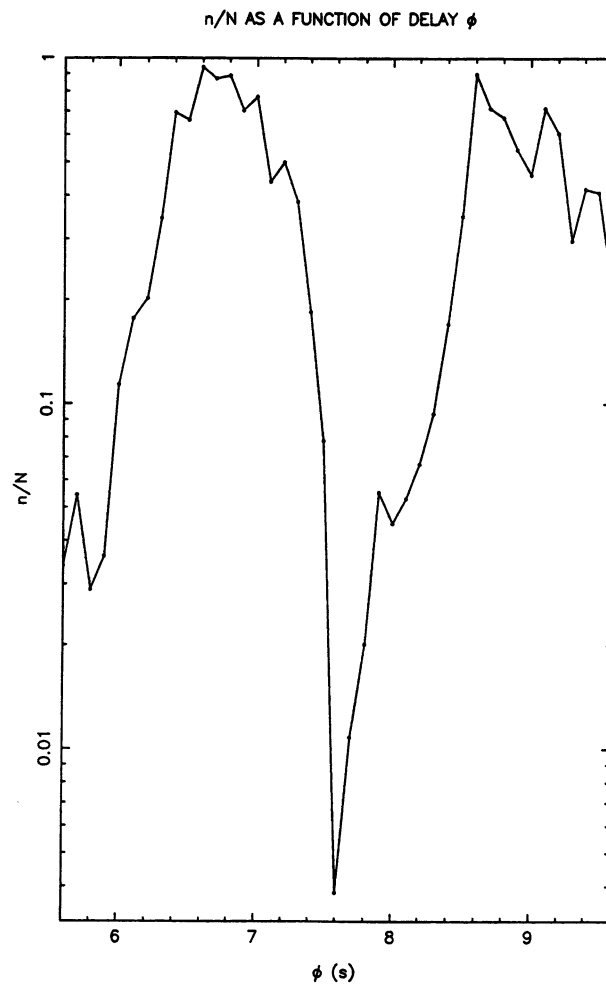


Figure 2.8: Net-excitation analysis of set 327. Although the threshold-correlation method gives as strong a correlation here as for set 55, the net excitation analysis does not show nearly as dramatic a dip as in Fig. 2.6

We can use this figure to estimate the realistic chance probability of the threshold coincidence correlations in the RTM data as follows. The first 20 bins in Fig. 2.11 contain 21 data sets. This suggests that the realistic probability is that any one bin will contain one data set in each  $10^4$  trials. Since the width of each bin is  $\Delta q = 10^{-6}$ , the true probability that a data set will give  $q$  less than  $10^{-6}$  (i.e., will fall in the first bin) is

$$p(q < 10^{-6}) \approx 10^{-4}. \quad (2.12)$$

Thus, a more realistic estimate of the a priori probability that the two hours of data which RTM analyze will show the sort of threshold coincidence correlation they find is  $10^{-4}$ . This estimate does not, of course, allow for other effects, such as the selection of the data set and

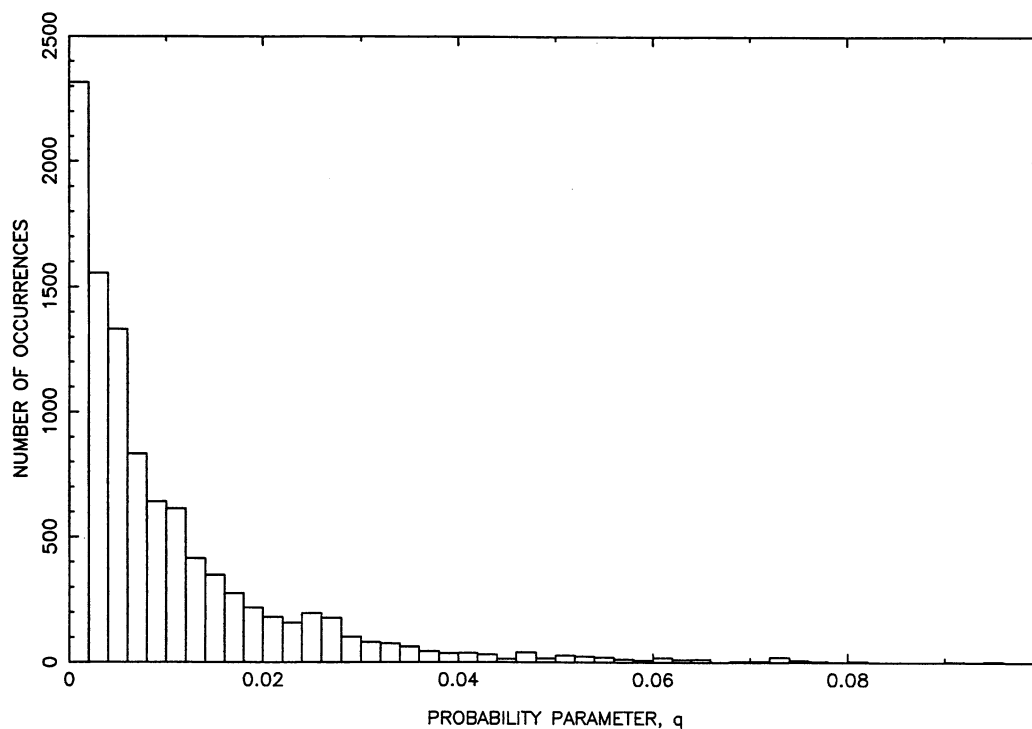


Figure 2.9: Relative frequency distribution of the values of the parameter  $q$  in our second simulation run. This parameter is used by RTM as their probability estimate. If this were the true probability, this figure would be a straight line through the origin.

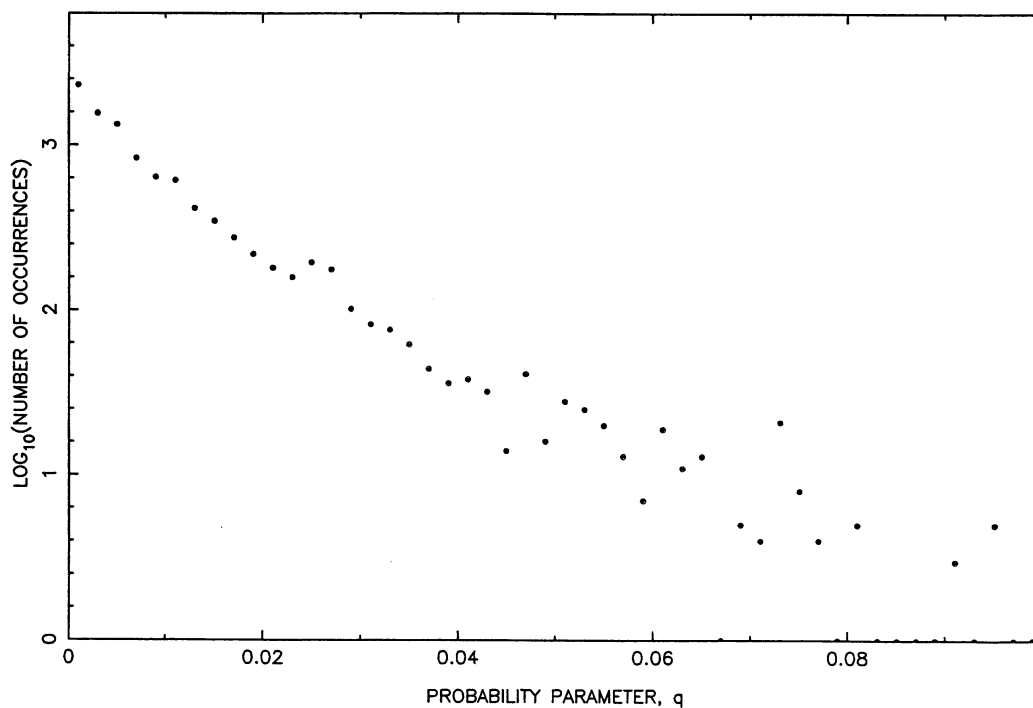


Figure 2.10: Logarithm of the previous figure, showing a nearly exponential distribution.

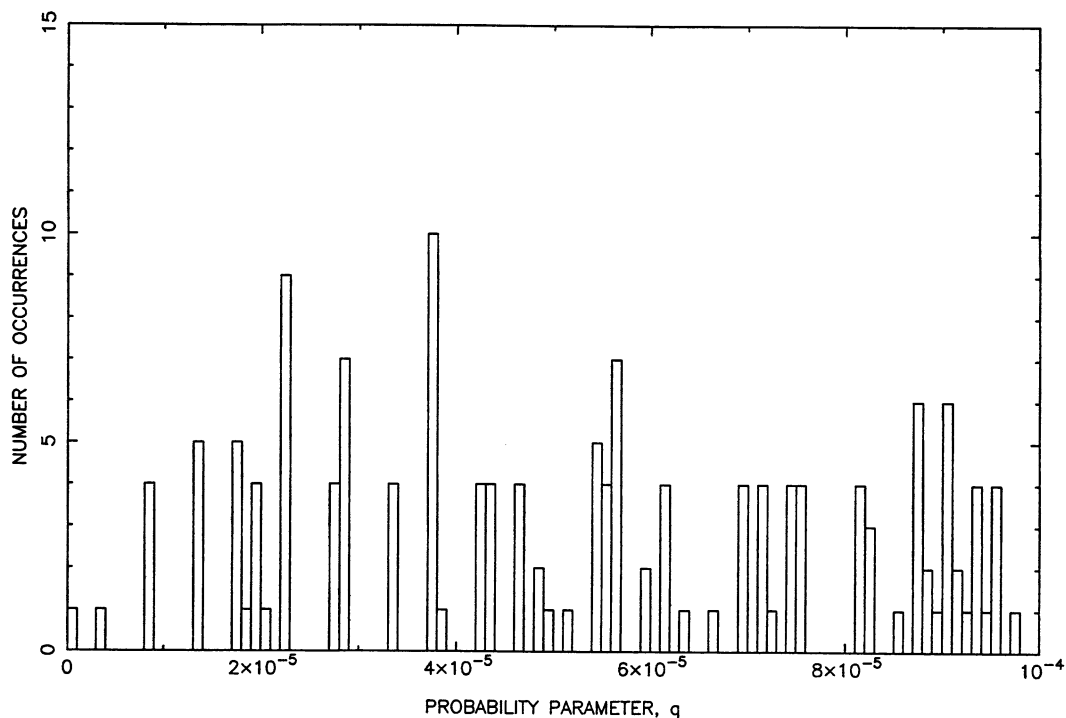


Figure 2.11: Frequency distribution of  $q$  for small  $q$ , allowing an estimate of the distribution of unlikely correlations. If the distribution in the previous figure is fit by a straight line for small  $q$ , then its slope in this figure would be nearly horizontal because of the greatly enlarged scale for  $q$ .

the *a posteriori* nature of the analysis method. In the next section we take these less easily quantified effects into account. We shall also show that it is possible to find evidence within the RTM threshold-coincidence analysis itself that our simulation probabilities are closer to the true probabilities than RTM's own estimates.

## 2.5 Reassessment of RTM correlations

We shall now make a full reassessment of the probabilities of the correlations RTM have found, in the light of our Monte Carlo simulations. We shall study the results of five papers, all of which reported coincident events: two (Amaldi *et al.* 1987; Aglietta *et al.* 1989) found coincidences between the gravitational wave detectors and the Mt. Blanc neutrino detector (see Section 2.5.1); two (Amaldi *et al.* 1989; Aglietta *et al.* 1991b) found coincidences between the gravitational wave detectors and the Kamiokande particle detector (see



Section 2.5.2); and one (Aglietta *et al.* 1991a) found coincidences between gravitational waves and IMB (see Section 2.5.3). (Another paper (Amaldi *et al.* 1988) found correlations between the two gravitational wave detectors themselves, but the probabilities found were not so unusual, so we review it briefly in the Appendix 2.7.2.)

We shall deal in this section with the main analysis methods which RTM use; though in the interests of completeness, we have included many of the details of the various analysis papers in the appendix at the end of the paper. We shall first reassess the Mt. Blanc neutrino–gravitational wave coincidences; then we shall re-examine the Kamiokande– and IMB–gravitational wave coincidences.

### 2.5.1 Reassessment of Mt. Blanc–gravitational wave coincidences

The RTM calculations of probability (Aglietta *et al.* 1989) are seriously affected by certain *a posteriori* choices they have made. Using our simulations in Section 2.4, we have already assessed the effects of some of these choices—delay time and thresholds—on the results of their threshold-coincidence analysis, coming to the conclusion that the correlations they find have an *a priori* probability of about  $10^{-4}$  in any single random data set. We have also shown that probabilities derived from the net-excitation method are not reliable below values of a few times  $10^{-4}$ .

In this section, we firstly examine the behaviour of RTM’s data set without the 50 s of data which they used to choose their first time delay of 1.4 s, and we attempt to use this data set to test the predictive power of their choice. This gives an estimate of the probability of the RTM correlations which is independent of one’s guess as to how much freedom RTM had to adjust their delay parameter.

We then consider other significant *a posteriori* choices that RTM made that made it easier for them to obtain correlations. We shall see that some are quantifiable, while the effects of others can only be guessed at. The overall effect of these considerations is further to increase the likelihood of RTM’s discovering the correlations which they find.

#### Contamination of 2-hour data by including the “eyeballed” set.

An important issue is the fact that RTM included in their full data sets the original 50 s stretch of data that contained the  $5\nu$  burst that originally suggested to them that they should search for a time-delay of about 1.4 s. RTM are aware that this biases their probabilities

and at one point attempt to show that this has a negligible effect on the final result. We will explain below why their argument is wrong. We will then show how removal of the 50 s of data can be used to control for RTM's ability to choose the time delay, by assessing the predictive power of a 1.4 s delay chosen from those 50 s of data, used for the whole data set excluding those 50 s.

**Effect of contamination.** It is straightforward to estimate the effect of this contamination on the threshold coincidence method that RTM apply to the Mt. Blanc data. In the 2-hour stretch they analyze, they find 13 coincidences at the adopted threshold. Against an expected value of 2.29, Eq. 2.7 gives a chance probability of about  $10^{-6}$ . If we exclude the first "neutrino" of the Mt. Blanc burst, which is clearly in coincidence with gravitational waves in summation above the threshold of 150 K (and is the only one), then the number of coincidences at this threshold falls to 12. This gives a chance probability of about  $5 \times 10^{-6}$ . This is before corrections for the arbitrariness of the threshold, time-delay, etc. The contamination thus makes their threshold-coincidence probabilities a full factor of 5 too small.

The contamination is much greater in the net-excitation method. Consider the statistic  $\sum_{i=1}^{N_\nu} E(t_i + \phi)$  in Eq. 2.1 which seems to give such an unusually large value. The five "neutrinos" originally "eyeballed" in Fig. 2.2, with the delay deliberately chosen so that the Rome gravitational waves so delayed with respect to the "neutrinos" are appreciably higher than the average background, will each add about  $(82.4 - 28.6 \approx) 55$  K extra to this sum (see Fig. 2.4) at delays of both 1.4 s and 1.1 s. This artificially increases the sum by about 275 K and so, when divided by 96 for the number of "neutrinos" detected in the two hours under analysis, this contributes about 3 K to  $C_+(1.1)$  and  $C_+(1.4)$  (see equation Eq. 2.2). This would considerably alter the ranking order of  $C(1.1)$ . Fig. 12 of Aglietta *et al.* (1989) shows that if the value of  $C(1.1)$  were 3 K less, there would be about 20 "background" values greater than  $C(1.1)$ , while there were none before. That is, without the "eyeballed" data, the RTM estimate of the probability of the correlations in the rest of the 2-hour data would be raised to  $2 \times 10^{-5}$ . This factor of 20 still leaves the probability below the range of reliability of the net-excitation method.

**Contamination correction as a way of controlling for the time-delay freedom.** Excluding the “5-neutrino burst” is in fact another way of compensating for the freedom to choose the time delay in the correlation analyses. If we allow only the original RTM “eyeballed” time-delay of 1.4 s and exclude that data set from the subsequent analysis, we would obtain an unbiased result that tests the ability of the original set to predict correlations in the extended data set. For the net-excitation method, this would remove the principal degree of freedom. However, all we have been able to do is perform that test for the revised delay of 1.1 s, where we found the probability went up by a factor of 20. We should really apply this correction to the original time-delay of 1.4 s, but the RTM papers do not provide enough information for us to be able to do this. However, we can be certain that the proper correction would raise the probability even further, since in the full data set a delay of 1.1 s gave a better correlation than did 1.4 s; while in the data that one removes (containing the 5  $\nu$  burst), the 1.4 s time-delay was better (see Section 2.4.3).

The threshold-coincidence method is, of course, also contaminated by this, and we have seen that this correction is a factor of 5. This is a correction only for the freedom to choose time-delays, not for the threshold freedom. Since in our simulations (which are not affected by this contamination because we do not look at the first 50 s to get a time-delay and then re-use this stretch of data in estimating the probability that the full set shows a correlation) we took a correction factor of 15 for time-delays (a 15 s span rather than RTM’s 1 s), the factor of 5 takes the corrected RTM probability most of the way toward our simulation estimate. Moreover, the remarks in the last paragraph about using the original time-delay of 1.4 s apply here too. This will raise the correction still closer to (if not beyond) our factor of 15. *We find, therefore, that the contamination effect can be used to control for the time-delay freedom, and when one does so one finds consistency with the probabilities of  $10^{-4}$  produced by our simulations.*

**Problems with the RTM contamination correction.** RTM realized that the contamination of the 2-hour data by the “eyeballed” data was a problem, and they attempt to show that it does not really change things by calculating the net-excitation ranking statistic (comparison of  $C_+(1.1)$  with random “background” values) with and without the 5 “neutrinos” of the Mt. Blanc burst. They find no significant change (Fig. 5 of Aglietta *et al.* 1989). However, the comparison is flawed because they used only  $N = 10^3$  background values to

calculate the “probability” of the correlation, both with and without the  $5\nu$  burst. Such a calculation can (according to our argument on the independence of the background values) indeed distinguish between data sets that have a chance probability greater than about  $10^{-3}$ , but unfortunately RTM adduce this calculation as evidence that a data set with a probability of  $10^{-6}$  is uncontaminated. Even if their method were reliable, they would have had to have used at least  $10^6$  background points to have drawn any conclusions.

RTM tell us (Pizzella & Pallottino 1991) that they have, in response to our criticism, subsequently performed such an analysis with  $10^6$  points and find that their net-excitation probability goes up by a factor of 5 when the original “neutrinos” are excluded. While this takes them some of the way toward the  $10^{-4}$  level that we feel the correlations really warrant, they still have not compensated for changing from 1.4 s to 1.1 s, and they are, in any case, using a method whose probabilities are unreliable at this level.

#### Further corrections to the probability of the Mt. Blanc correlations

We have shown from our simulations that the correlations in the Mt. Blanc data occur with probability  $\sim 10^{-4}$ . We have confirmed this by removing the data from which the 1.4 s delay was chosen, and testing the predictive power of this delay on the rest of the data. We start this section, therefore, with the estimate that, for the given Rome-Maryland-Mt. Blanc data set, the probability that RTM would have found the correlations they did find is about  $10^{-4}$ .

**Selection of the data set to analyze.** Through our simulations and our attempts to correct for time-delay and threshold freedom in the RTM analyses, we have arrived at the conclusion that the given Rome-Maryland-Mt. Blanc data set contains correlations with a real probability of about  $10^{-4}$ . While larger than the RTM claim of  $10^{-6}$ , this is still potentially significant. However, we now have to turn to a number of corrections that have to do with other *a posteriori* choices made by RTM.

The first is that RTM see their correlations only in a particular 2-hour stretch of data, which was not selected because of any property of the 50 s “eyeballed” data set. Indeed, RTM looked for correlations in other, earlier, data sets and found none at the same time delay of about 1.2 s. Also, they examine longer and shorter data sets and find that the effect becomes much weaker for periods less than about 50 minutes and greater than about 150 minutes (Fig. 9 of Aglietta *et al.* 1989). Indeed, they seem to regard this as evidence for

the reality of their correlations, since if they were associated with the supernova, then one would expect them to be transient.

However, when assessing the significance of correlations, one must be careful to start from the *null hypothesis*, that the correlations arise by chance. Then it is clear that one's ability to choose the data set in which one finds correlations is another free parameter, like the time-delay itself. Since one has no *a priori* idea of the length of the period during which these correlated "neutrinos" and gravitational waves (or new particles) should have been emitted by the supernova, it is fair to expect that if correlations as strong as the ones RTM found had appeared instead in, say, a longer or shorter stretch of data, RTM would have treated them just as seriously.

In fact there are *two* variable parameters here: the length of the data set and its starting time. RTM make a natural *a priori* choice in selecting a data set which includes the Mont Blanc  $5\nu$  burst, but it need not have been 2 hours long and it need not have been *centered* on the burst. It would have also been natural to have looked for phenomena either immediately preceding the putative collapse event or immediately following it; indeed, on physical grounds it seems rather unlikely that any correlated phenomena would have occurred *both* before and after the collapse, since the physical conditions are so different on either side.

The length of the data set is even more important. RTM analyze a 2 hour data set, but again give no physical reason for having made this choice *a priori*. The reason for this choice seems to come from Fig. 9 of Aglietta *et al.* (1989), where, for a fixed delay of 1.2 s (given by "eyeballing"), RTM compare  $C_+(\phi)$  and  $C_+(\delta_1, \delta_2)$  for different values of the length of the data set, and show that the best correlations for the net-excitation algorithm occur for lengths between about 100 and 130 minutes, with the "probability" of the correlation increasing fairly sharply by about 1 or 2 orders of magnitude outside a window from about 70 to 150 minutes. RTM apparently used this information to select the data set they analyzed. In fact, RTM stress that the 2 hour length of data is not optimal: 135 minutes is better. But it is clear that even so they have made a considerable optimization by choosing a value near the "best" one, when they could have chosen a length of anything from, say, a few minutes to 36 hours.

We have not attempted to simulate this freedom to choose the data set in our Monte Carlo analysis; it would have been computationally very expensive. We also do not know from the published papers how many data sets RTM actually looked at. In the absence of

simulations, the following argument gives us some idea of the size of the effect.

We would like to know how many essentially independent data sets RTM could have analyzed. Let each data set contain the Mt. Blanc burst, and let us take a minimum reasonable data set length  $L$  to be 8–10 min. If we enlarge the set by a factor of 2, the larger set will have statistics reasonably independent of those of the smaller included in it. Each such doubling of the length produces a new “independent” set, until  $L$  reaches 36 hours, the total of the data apparently available to RTM initially. This requires 8 doublings, giving 9 sets. For the shorter sets there are actually two independent sets, one ending with the Mt. Blanc burst and the other beginning with it. Doubling these “post-Mt. Blanc” sets until the Maryland detector goes off line because of its electrical problems adds 5 more sets, giving 14 in all.

We shall therefore take a factor of 10 to be a reasonable lower limit on the correction we need to make for this selection effect. *This raises our estimate of the probability that RTM’s analysis methods would have found correlations in entirely random data to about  $10^{-3}$ .* Next we turn to the problem that their analysis methods were themselves invented *a posteriori*.

**“Trial and error” analysis.** Every textbook introduction to statistical analysis emphasizes the problem that, the more often one analyses a given set of random data in different ways, the more likely it is that one will uncover a correlation of apparent significance. In our simulations in Section 2.4 we have accordingly reported *all* the trials we did. Unfortunately, it is impossible from RTM’s papers to learn whether they performed other analyses of the data that they do not report. We have indicated at several places in this paper our guess that they may (or even should) have done so.

For example, the most natural kind of analysis to have done with two gravitational wave streams and the Mt. Blanc data is a triple-coincidence analysis, where one identifies gravitational wave “events” by setting a threshold separately on these two data streams. The threshold need not be arbitrary: a reasonable one is a level where one expects only a few coincidences over the selected data set if the data are random (a low “false alarm rate”). RTM do not report such an analysis. Instead they report a double coincidence analysis in which the gravitational wave data are added together before being thresholded, and they search many thresholds. They also introduce a non-standard method, the net-excitation method. However, as we shall see, RTM *do* report having done such a triple-coincidence

analysis for the Kamiokande data and gravitational wave detectors.

Having used certain methods for the Mt. Blanc data, they then do not stay exclusively with them for the KII data. The net-excitation analysis is done but not examined in detail. The threshold-coincidence method is not reported, but the results of the triple coincidence method are. And the length of the data set is changed. The papers do not tell us if RTM performed, say, the threshold-coincidence analysis of the KII data over the original time-span and did not report it because the results were not very significant.

One cannot argue that these tests are all roughly equivalent, so that if a correlation shows up in it will show up in all: this is not necessarily the case. For example, our Monte Carlo simulations produced two “good” correlations as measured by the threshold-coincidence method, but one of them gave a good correlation using the net-excitation method and the other did not. Here, the choice of the analysis technique used makes a difference of a factor of  $10^3$  in the “probability” obtained. These methods all measure different things (though some methods are partly dependent on each other in ways which are not clear). So the significance of a reported correlation is diminished if other tests were applied that gave null or insignificant results, simply because the other tests *could* have given correlations (even if they did not).

Another worrying aspect of this is that there are occasions where it appears that a secondary analysis was designed after a primary analysis, and may therefore have been guided by the results. An example of this, which we have already seen, occurs in the design of the net-excitation method in Aglietta *et al.* (1989). When calculating the “background” neutrino–gravitational excitation to compare with the measured value given by Eq. 2.1, RTM make an unexpected choice: they use Eq. 2.2, in which the gravitational wave data streams are taken at different times, rather than simply shifting both gravitational wave data streams by the same amounts.

While this would not be unreasonable as an *a priori* choice (provided they had used enough data to ensure independence of the background values), the problem is that RTM by this time appear already to have performed the Maryland-Rome correlation analysis (Amaldi *et al.* 1988), which showed that the two gravitational wave detectors had an unusually high number of coincidences, at zero relative time delay, during the period under analysis. RTM should have known that by calculating the background as they have, they have obtained a marginally lower value for the apparent probability than if they had kept the two gravitational wave data streams tied together.