Reassessment of the reported correlations between gravitational waves and neutrinos associated with SN 1987A

C. A. Dickson and Bernard F. Schutz
Department of Physics and Astronomy, University of Wales College of Cardiff, Cardiff, Wales, United Kingdom
(Received 6 September 1994)

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-h period near the explosion of Supernova 1987A. 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 simulated pseudorandom 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.

PACS number(s): 04.80.Nn, 95.55.Vj, 95.85.Sz, 97.60.Bw

I. INTRODUCTION

At about the time of the Supernova explosion SN 1987A there were, unfortunately, only two gravitational wave detectors in operation [1]. 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 [2], IMB [3], and Baksan [4] — registered a coincident burst. Unfortunately, only one gravitational wave detector was recording data at that time (Rome), and that was affected by seismic noise [1]. However, at the time of a somewhat earlier "neutrino" burst in the Mont Blanc detector [5], 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 (RTM) Collaboration [6], found (i) that with a delay of 1.4 s with respect to the five 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 [7], with a chance probability of 3%), and (ii) there was a modest correlation between the Rome and Maryland gravitational wave detectors in a 7-h period around the time of the Mont Blanc burst [1] (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 SN 1987A had been detected.

However, in subsequent analyses, RTM searched a larger stretch of data for further events such as those reported earlier [7], where a gravitational wave detector is excited a fixed time before a particle is detected. This has led to a series of papers [7–11] 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 [7,8] and between the gravitational wave detectors and the Kamiokande [9,10], Baksan [10], and IMB [11] 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-h period. If they are due to neutrinos and gravitational waves from SN 1987A, the energy involved would be huge, many thousands of solar rest masses converted into gravitational wave energy *for each event* [12]. 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 [13]. 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 [9], Baksan [10], and IMB [11] 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 [10]. 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 Sec. II 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 Sec. III 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 Sec. IV 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 Sec. V 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 analyzed, 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: all analyses of a data set, whether or not they give the results expected or desired by the analyzers, should be stated; data sets should be carefully examined individually and the results reported before they are combined; 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; 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 data set, any new data should be analyzed in the same way as the original data were.

II. THE GRAVITATIONAL WAVE AND NEUTRINO OBSERVATIONS

The observations of SN 1987A are well documented [14], 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.

A. Particle observations

There were four particle detectors in operation during the relevant period: Mont Blanc (variously called UNO or LSD) [5], Kamiokande (K II) [2], IMB [3], and Baksan [4]. All four were in operation during the whole of 22–23 February 1937. The optical brightening of the Supernova took place between about 2 h and 11 h UT on 23 February. Neutrinos would have been expected at any time up to 24 before this, allowing time for the hydrodynamic shock to reach the star's surface and cause the optical display.

At about 2 h 52 m 37 s UT, Mt. Blanc observed a burst of five "neutrino" events [5]. This burst had a probability about 2×10^{-3} of arising purely from the Poisson background during a period of 24 h immediately preceding the observation of the optical Supernova event [5]. 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 [14].

The later burst, however, at about 7 h 35 m 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 [2] detected 11 neutrinos at 7 h 35 m 35 s UT (± 60 s) within a time interval of 13 s, with energies between 7.5 and 36 MeV. IMB [3] reported eight neutrinos at 7 h 35 m 41 s UT during an interval of 6 s, with energies from 20 to 40 MeV. Baksan [4] detected five neutrinos at 7 h 36 m 11 s 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 7 h 36 m 00.5 s UT and 7 h 36 m 18.9 s, discovered in the off-line analysis [15]. 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 [2] had an absolute timing uncertainty, Δt_K, of ± 60 s; while the absolute uncertainty Δt_B in the Baksan clock [4] 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 [5] accurate to ± 2 ms and IMB [3] to ± 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 Secs. III and V.

B. Gravitational wave detectors

The Rome and Maryland room temperature bar gravitational wave detectors operated satisfactorily at least from 18 h 24 m 3 s of 21 Feb 1987 to 6 h 2 m 3 s of 23 Feb 1987, a period that includes the Mt. Blanc burst, but excludes the time of the KII-IMB-Baksan events. Soon after 6 h on 23 February, the Maryland detector experienced electrical problems; and at 7 h 35 m UT, the time of the KII-IMB-Baksan coincident burst, there were seismic disturbances in Rome. RTM confine all their analyzes to the period before 6 h 2 m 3 s 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 three 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 6 h, 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 [8]. 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 ± 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 ± 0.1) s to obtain the true time.

III. SUMMARY OF THE MAIN RTM ANALYSIS METHODS

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

A. The main RTM analysis methods

1. *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 antennas $E_R(t)$ 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:

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

where ϕ is a chosen offset time, t_i is the arrival time of the ith neutrino, N_ν 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 ϕ 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 [8–11] use the sum algorithm only.) The decision of RTM [8] 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:

$$C_*(\delta_1, \delta_2) = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} [E_R(t_i + \delta_1) * E_M(t_i + \delta_2)] , \quad (2)$$

where δ_1 and δ_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 ϕ by defining

$$n(\phi) = \text{count}_{\delta_1, \delta_2} [C_*(\delta_1, \delta_2) \geq C_*(\phi)] , \quad (3)$$

where "$\text{count}_{\text{range}}[\langle\text{condition}\rangle]$" 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 δ_1 and δ_2 always includes ϕ, 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

$$p(\phi) = n(\phi)/N_b. \quad (4)$$

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 Sec. III A 2. (We discuss the effect of small-number statistics $[n(\phi) \sim 1]$ in Appendix A 1 a, and we return to the question of the distributions of $C_*(\phi)$ and $C_*(\delta_1, \delta_2)$ in Appendix A 2 b.)

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 ϕ may give values of C_* that are very close, but they could be far apart in $n(\phi)$.

2. 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 δ_i such that the background value is taken from the same period as the signal $C(\phi)$ [8]. This is to avoid problems due to possible nonstationarity of the noise. Now, in a 2-h stretch of data, where RTM find their strongest correlations [8], there are 7200 1-s samples from each gravitational wave detector. On the null hypothesis (no genuine correlation), there are thus about 1.4×10^4 independent random numbers in the original data. These numbers are combined in various ways using Eq. (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. (4) for the smallest apparent probabilities.

3. 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

$$E_*(t) = E_R(t) * E_M(t) \tag{5}$$

(where again $*$ is $+$ or $\times$), 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 ϕ 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_ν neutrinos and $N_{\rm GW}$ gravitational wave events randomly (uniformly) distributed, the expected number of coincidences is

$$\bar{n} = \frac{N_\nu N_{\rm GW}}{N_t}. \tag{6}$$

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

$$p_{\bar{n}}(n) = \sum_{r=n}^{\infty} \frac{\bar{n}^r e^{-\bar{n}}}{r!} = 1 - \sum_{r=0}^{n-1} \frac{\bar{n}^r e^{-\bar{n}}}{r!}. \tag{7}$$

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.

IV. MONTE CARLO SIMULATIONS

A. 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 pseudorandom 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 pseudorandom 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.

B. Properties of the pseudorandom 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-h 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 [8]).

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 [8]. (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* [16] uniform random number algorithm RAN1.

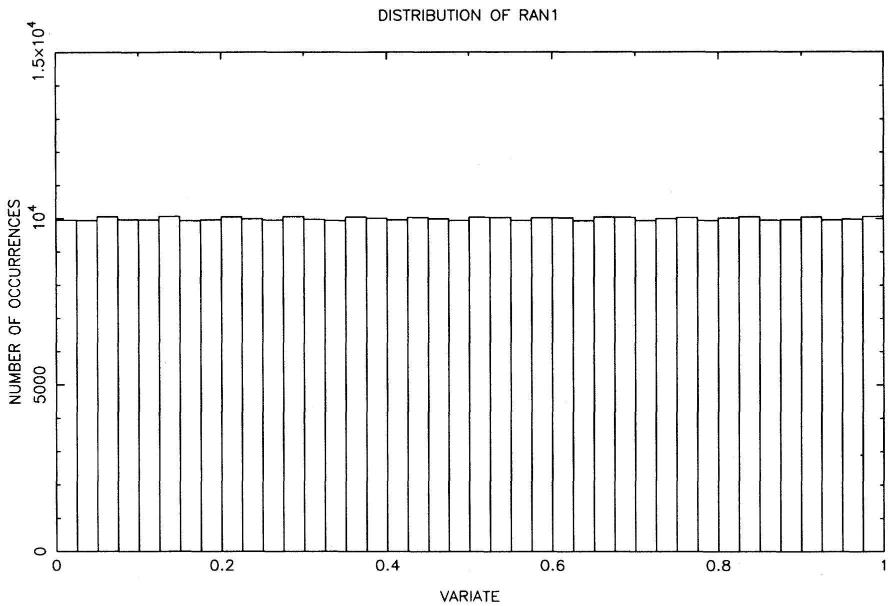

FIG. 1. Distribution of the pseudorandom number generator.

The cycle length of this random number generator is said to be infinite for all practical purposes [16]. We demonstrate its distribution by generating and binning the first 4×10^5 numbers in Fig. 1.

C. Method of analysis of the pseudorandom data

1. 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.

2. Selection of a threshold

In choosing the threshold T, we are guided by what RTM say about their choice [8]. 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 Ref. [8], then the "true" probability of a correlation would be larger.

3. 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 analyzing the Mt. Blanc data, RTM changed their "best" delay from 1.4 s [7] to 1.2 s [8] and then to 1.1 s [8], 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 optimize 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 Ref. [8], this optimization 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 ν burst. RTM also tell us that they would never have adjusted the delay by more than about 0.5 s–1 s [17], 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 data sets, 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 $1.4 \pm$ (say) 0.5 s when the phenomenon, by hypothesis, occurs over a period of 2 h 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 s to $+0.8$ s, which was not necessary for the calculation of the strength of the correlation at -1.4 s, which they had postulated. It was this search that led to their later adoption of a delay of 1.1 s. Also, using the threshold coincidence method, RTM searched from -50 s to $+50$ s, for a fixed threshold, and found no correlations stronger than those around 1.2 s. This is not surprising since the threshold was optimized for the chosen delay of 1.2 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 s delay.

(a) *On the initial selection of the 1.4 s time delay.* RTM initially inspected a small stretch of data containing the 5 Mt. Blanc neutrinos (see our Fig. 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" [7]. Since this criterion is just an "eyeball" implementation of their own net excitation method [Eq. (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. 3 we plot the statistic

$$C(\phi) = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} E_R(t_i + \phi), \tag{8}$$

which is the single-detector version of C_* of Eq. (1). Our figure contains two plots: (a) uses all six neutrinos that are shown in Fig. 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 [7,8]. 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. 1] applied to all the data of Fig. 2. Our results are shown in Fig. 4. Here, the picture is very different: there is little to choose between time delays near 1.5 s and those near 5.5 s and 8.5 s. In fact, if one uses only the 5

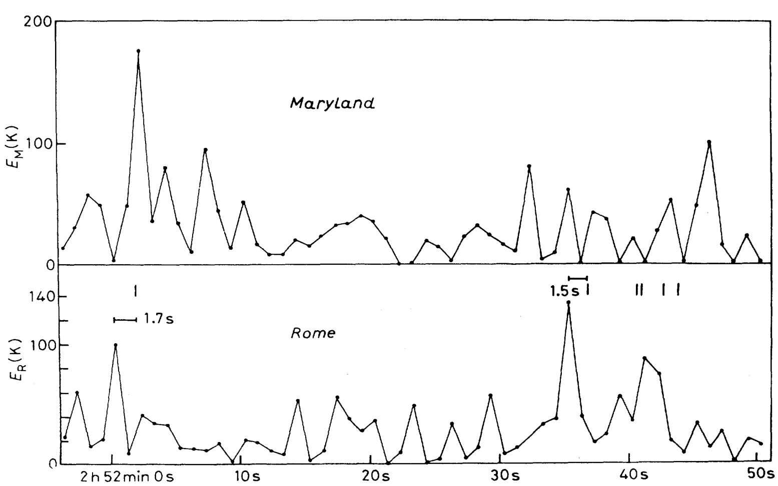

FIG. 2. First indications of the Mt. Blanc correlations. RTM originally had only the Mt. Blanc neutrinos and the Rome data from which to select a delay of 1.4 s. The Maryland data were obtained later, and appear in the analyses in Ref. [1]. (Reproduced from Ref. [8] with permission.)

"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 Sec. I), 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.

(b) *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 predetermined range) for a given set of random data.

We fix the range of available time delays by staying with the RTM model [7] of ascribing the delay to the effect of a neutrino mass, m_ν. The time delay between

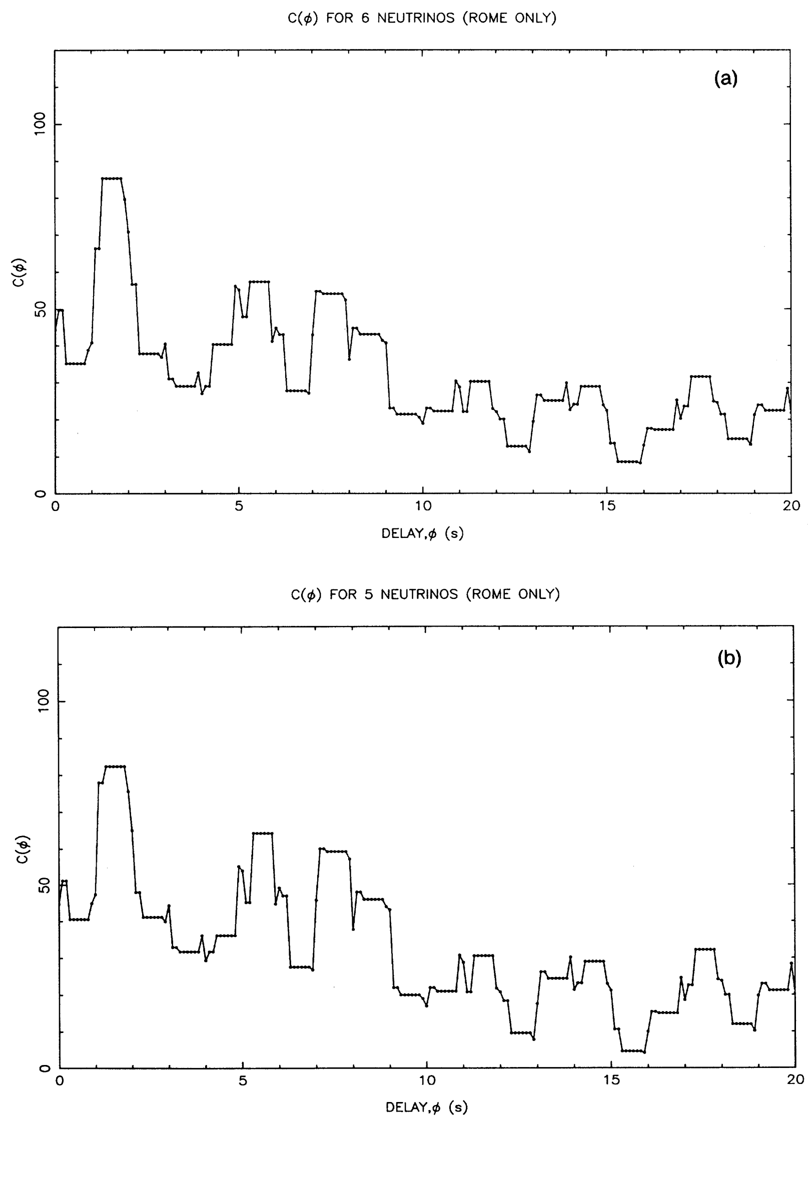

FIG. 3. Searching for good time delays using the net-excitation method applied to the data set of Fig. 2, using Rome data only. (a) contains all six neutrinos seen in Fig. 2, while (b) omits the isolated neutrino event near 2 h 52 m 2 s.

a gravitational wave traveling at the speed of light and such a neutrino with energy E_ν after traveling a distance d is

$$\delta t = \left(\frac{m_\nu c^2}{E_\nu} \right)^2 \frac{d}{2c}. \tag{9}$$

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 [7] 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

$$0.0 \text{ s} \le \delta t \le 14.9 \text{ s}, \tag{10}$$

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

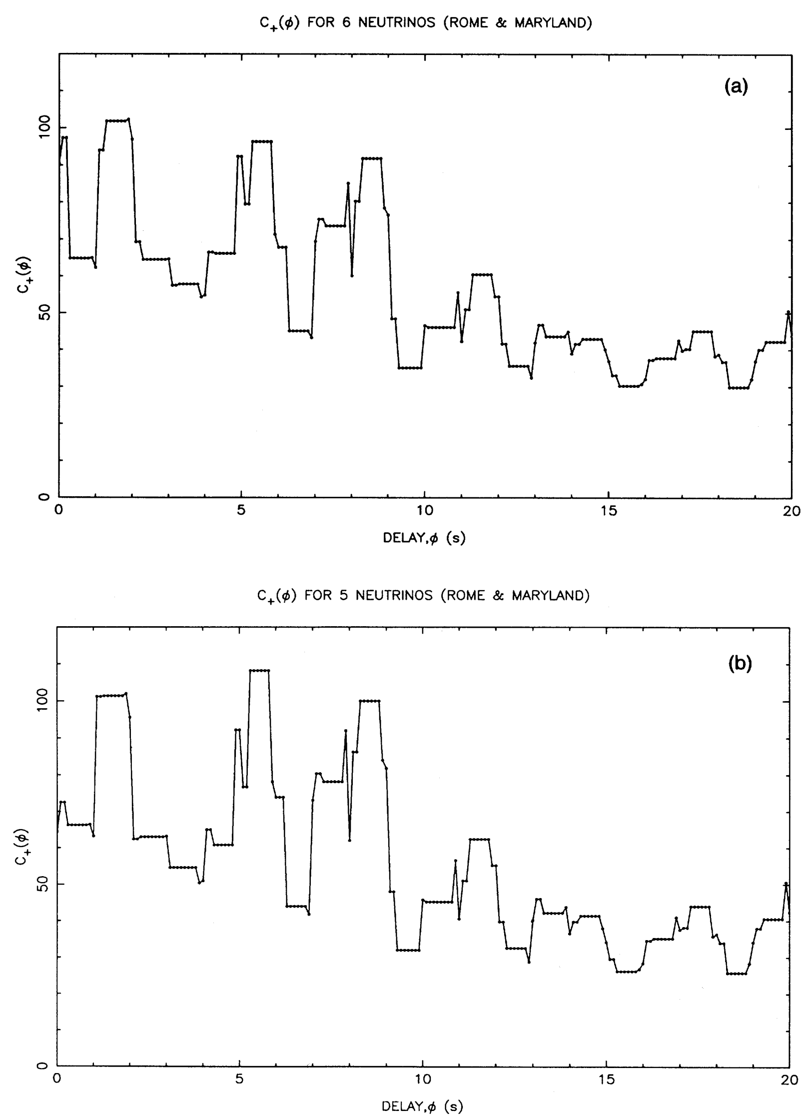

FIG. 4. Searching for good time delays using the net-excitation method applied to the data set of Fig. 2, using both Rome and Maryland data. (a) contains all six neutrinos seen in Fig. 2, while (b) omits the isolated neutrino event near 2 h 52 m 2 s.

all had not the peaks in the gravitational wave stream been fairly near the neutrinos in Fig. 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. (12) below, and (2) (RTM scenario) if one feels that RTM's original eyeball estimate was binding, ±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 Sec. V A 1 we attempt to calculate the *a priori* probability of the correlations in the 2-h 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.

4. 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 Sec. IV B. 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. (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.

D. 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 (subsection V A 2 b below), we report *both* of our analyses here separately. We have not performed any others.

1. 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. 5 we present these results in the same way as is done in Fig. 14 of Ref. [8].

For this data set, there were 86 simulated neutrinos within the two hours, and at threshold $T = 110$ K there were 512 "gravitational wave events," giving an expected number of coincidences $\bar{n} = 6.116$, by Eq. (6). The Poisson probability of obtaining 22 coincidences here is [from Eq. (7)]

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

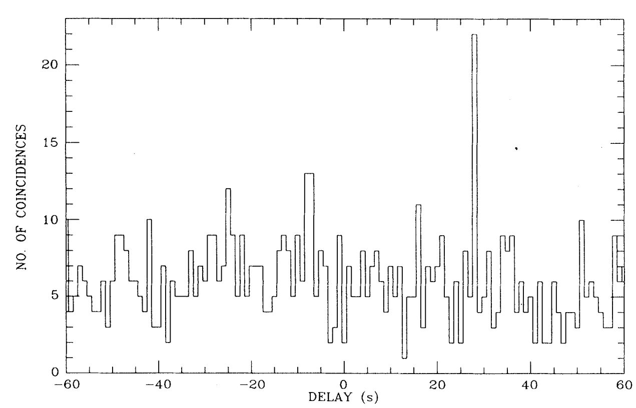

FIG. 5. Histogram of number of threshold-coincidences against delay time for data set 55 of the first simulation run.

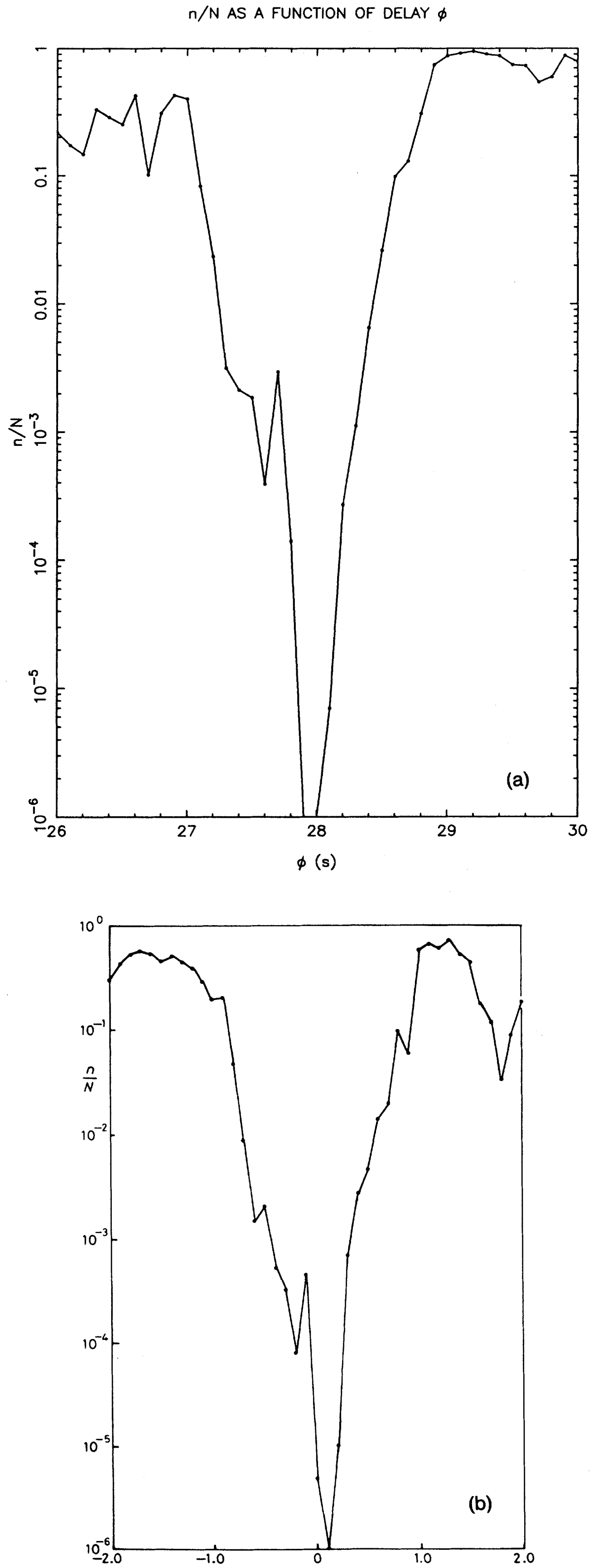

FIG. 6. Result of the net-excitation analysis of simulation set 55 (a) compared to the RTM analysis of the real neutrino and gravitational wave data (b). [(b) reproduced from Ref. [8] with permission.]

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 Ref. [8] appears in our Fig. 6(a). In Fig. 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. 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 data sets. 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.

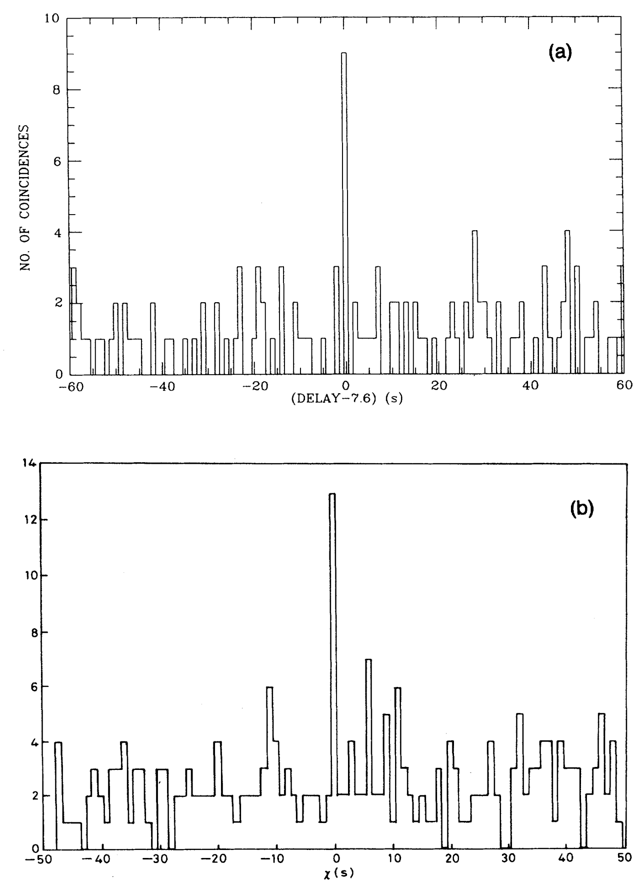

FIG. 7. Delay histogram for data set 327 of the second simulation (a) compared to the RTM histogram of the real data (b). [(b) reproduced from Ref. [8] with permission.]

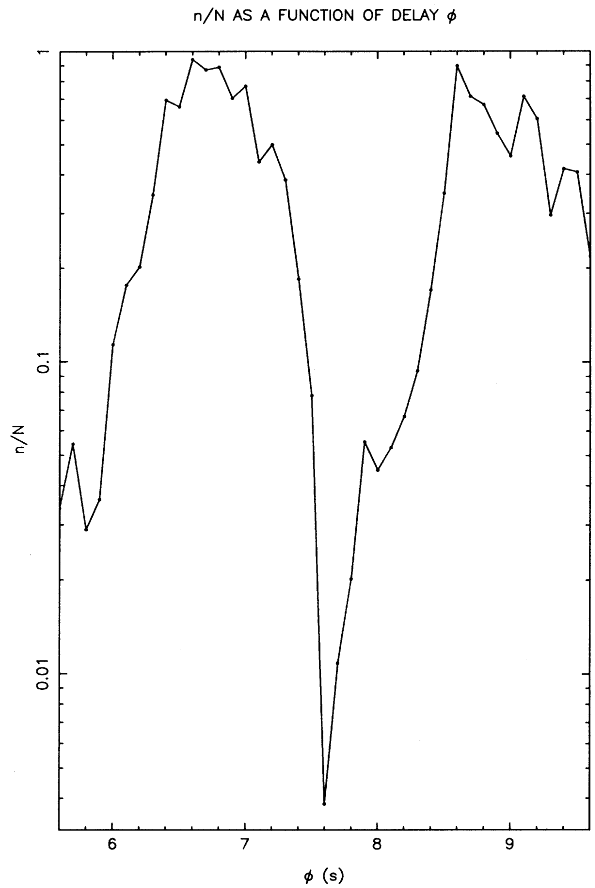

FIG. 8. Net-excitation analysis of set 327. Although the threshold-correlation method gives as strong a correlation here as for set 55, the net excitation analysis does not show nearly as dramatic a dip as in Fig. 6.

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.

2. 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.

a. The most improbable simulated data set. The most improbable data set in our second run was number 327, which had a peak of nine coincidences at delay 7.6 s at a threshold 170 K. The histogram of coincidences against time delays is in Fig. 7(a), plotted with the corresponding one for the RTM data (b). There were only 77 neutrinos in the 2 h of simulated data, and 83 gravitational wave events above this threshold. The number of expected coincidences is $\bar{n} = 0.888$ [Eq. (6)]. From Eq. (7), the peak of 9 has a probability of 4.5×10^{-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. 8 is by no means as dramatic as it was for the RTM data, or for our own Fig. 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

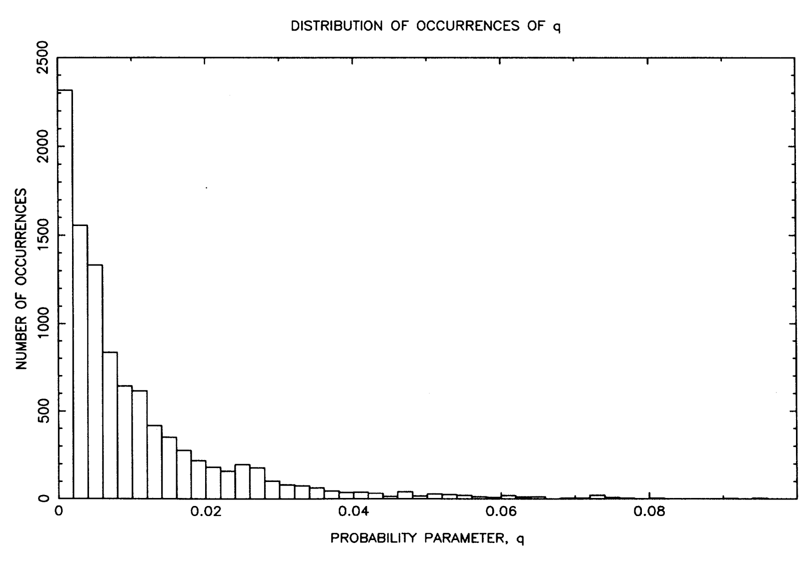

FIG. 9. Relative frequency distribution of the values of the parameter q in our second simulation run. This parameter is used by RTM as their probability estimate. If this were the true probability, this figure would be a straight line through the origin.

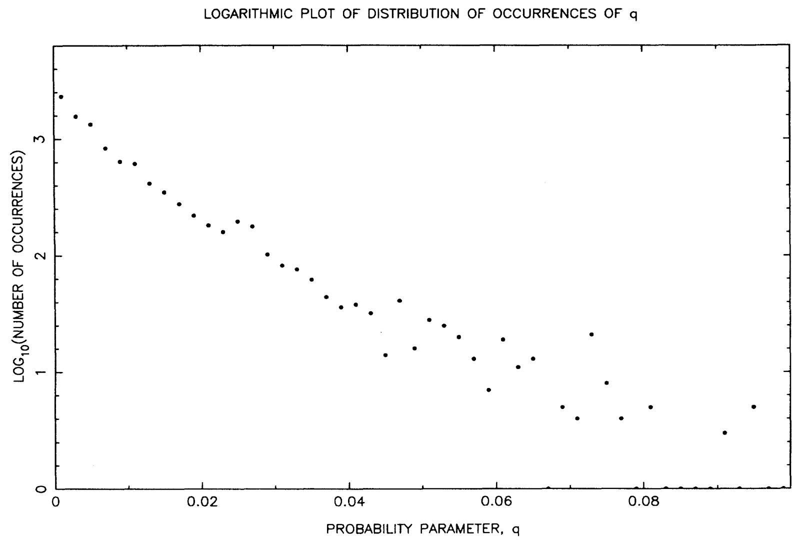

FIG. 10. Logarithm of the previous figure, showing a nearly exponential distribution.

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.

b. The relative frequency of occurrence of such correlations. Given the pseudorandom 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 ϕ 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:

$$q = \min_{\text{thresholds } T} \left\{ \min_{\text{time-delays } \phi} p_{\bar{n}(T)}[n(T,\phi)] \right\}, \tag{11}$$

where $p_{\bar{n}}(n)$ is given by Eq. (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. 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. 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. 11. Within the statistical fluctuations, the distribution is fairly flat, which is what we would expect if the behavior as $q \to 0$ is a regular extrapolation to zero of the low-q trend in Fig. 10, and does not become singular as $q \to 0$.

We can use this figure to estimate the realistic chance probability of the threshold coincidence correlations in

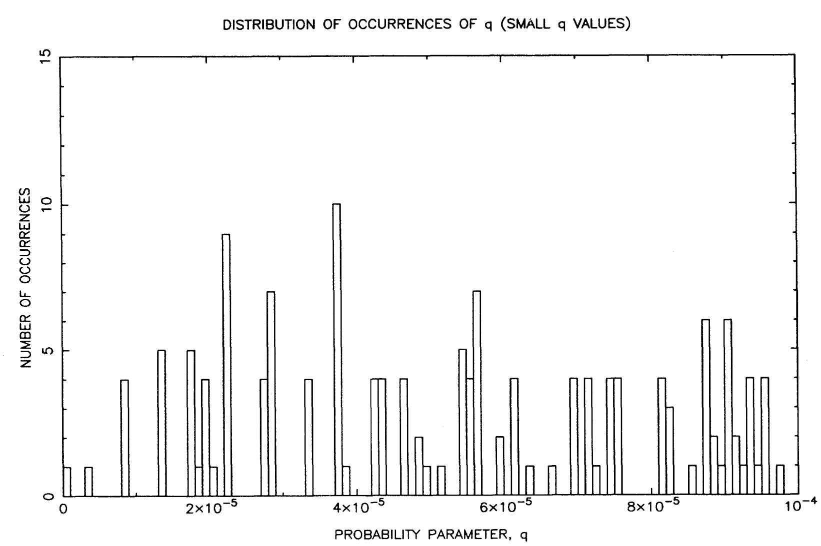

FIG. 11. Frequency distribution of q for small q, allowing an estimate of the distribution of unlikely correlations. If the distribution in the previous figure is fit by a straight line for small q, then its slope in this figure would be nearly horizontal because of the greatly enlarged scale for q.

the RTM data as follows. The first 20 bins in Fig. 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

$$p\left(q < 10^{-6}\right) \approx 10^{-4}. \quad (12)$$

Thus, a more realistic estimate of the a priori *probability that the 2 h 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 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.

V. 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 [7,8] found coincidences between the gravitational wave detectors and the Mt. Blanc neutrino detector (see Sec. V A); two [9,10] found coincidences between the gravitational wave detectors and the Kamiokande particle detector (see Sec. V B); and one [11] found coincidences between gravitational waves and IMB (see Sec. V C). (Another paper [1] 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 A 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 reexamine the Kamiokande– and IMB–gravitational-wave coincidences.

A. Reassessment of Mt. Blanc–gravitational wave coincidences

The RTM calculations of probability [8] are seriously affected by certain *a posteriori* choices they have made. Using our simulations in Sec. IV, 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 behavior 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.

1. Contamination of 2-h 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 ν 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.

(a) *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-h stretch they analyze, they find 13 coincidences at the adopted threshold. Against an expected value of 2.29, Eq. (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 5×10^{-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. (1) which seems to give such an unusually large value. The five neutrinos originally eyeballed in Fig. 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. 3) 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 2 h under analysis, this contributes about 3 K to $C_+(1.1)$ and $C_+(1.4)$ [see Eq. (2)]. This would considerably alter the ranking order of $C(1.1)$. Figure 12 of Ref. [8] 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-h data would be raised to 2×10^{-5}. This factor of 20 still leaves the probability below the range of reliability of the net-excitation method.

(b) *Contamination correction as a way of controlling for the time-delay freedom.* Excluding the "five-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 ν burst), the 1.4 s time delay was better (see Sec. IV C 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 reuse 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.*

(c) *Problems with the RTM contamination correction.* RTM realized that the contamination of the 2-h 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 five neutrinos of the Mt. Blanc burst. They find no significant change (Fig. 5 of Ref. [8]). 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 5ν 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 [17] 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.

2. 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}.

(a) *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-h 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 min and greater than about 150 min (Fig. 9 of Ref. [8]). 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 ν burst, but it need not have been 2 h 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-h 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 Ref. [8], 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 min, 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 min. RTM apparently used this information to select the data set they analyzed. In fact, RTM stress that the 2-h length of data is not optimal: 135 min 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 h.

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 h, the total of the data apparently available to RTM initially. This requires eight doublings, giving nine 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.*

(b) *"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 Sec. IV 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 nonstandard 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 tripl-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 one 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 Ref. [8]. When calculating the "background" neutrino–gravitational excitation to compare with the measured value given by Eq. (1) RTM make an unexpected choice: they use Eq. (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 [1], 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.

These points illustrate the difficulties that a posteriori invention of data analysis methods can create. We find these effects impossible to quantify, but inevitably they raise the probability of finding correlations.

3. Overall assessment of the probability

We now assemble our various corrections to arrive at an estimate of the chance that RTM would have found correlations of the level of significance that they have reported in entirely random data.

In the Introduction to Sec. V A we have put the chance probability of finding RTM-style correlations in a given random data set at about 10^{-4}. In Sec. V A 2 we have raised this to 10^{-3} because of their freedom to choose the data set. We cannot quantify the correction for what we have called "trial and error analysis," but it could be significant. One could conceivably get even another factor of 100 from this, since we actually found a variation of a factor of 1000 between different RTM analysis methods applied to the same simulated data set (see Sec. IV D 2).

We therefore conclude that, because of the considerable freedom that RTM had (and frequently exercised) in looking for correlations, the a priori chance probability of the correlations that they did find between the gravitational wave data and the Mt. Blanc neutrinos lies somewhere between 0.001 and 0.1.

B. Reassessment of KII–gravitational-wave coincidences

We turn now to the RTM coincidences between the Kamiokande detector and Mt. Blanc and the gravitational wave detectors. Of course if the correlations between the Mt. Blanc detector and the two gravitational wave antennas are due, even in part, to real neutrinos or other particles emanating from an astrophysical source, then similar correlations should be present in the data of other particle detectors, even though they did not exhibit obvious bursts of activity at this time, as did Mt. Blanc. RTM recognize that the correlations between the Mt. Blanc neutrino data and the gravitational wave data may have arisen by chance, so they rightly regard the acid test of the correlations to be their predictive power: do the particle data from Kamiokande show the same correlations with the same time delay? One would expect that the same methods of analysis as were previously used should yield correlations at a similar level of significance.

Unfortunately, RTM do not make a clean test of the predictions of the Mt. Blanc analysis; they do not simply apply the same analysis methods to the Kamiokande data as they used for the Mt. Blanc data. However, RTM do indeed claim to find correlations, with probabilities that lie between 10^{-3} and 10^{-8}, depending on the tests RTM apply. We shall reassess each of their methods in turn, but first we consider corrections that apply to all of them.

1. General corrections to the probabilities that RTM assign to KII

(a) *Clock correction.* As we mentioned before, the Kamiokande clock had an *absolute* timing uncertainty (difference from Universal Time) of ±1 min, although *relative* timings were very accurate. This could be rectified to a certain extent by demanding that the Kamiokande neutrino burst at 7 h 35 m was coincident with the burst in the IMB detector, whose clock was accurate. RTM try to find a clock shift that will bring the two bursts of neutrinos detected at 7 h 35 m 41 s (±5 ms) and 7 h 35 m 35 s ($\pm$ 60 s) UT in IMB and Kamiokande, respectively, into coincidence. But what does one mean by coincidence? Since the length of the Kamiokande burst [2] was 13 s and the length of the IMB burst [3] was 6 s, we have, unfortunately, a great deal of freedom in choosing when the two bursts should coincide. RTM use the criterion that the first neutrino of each burst should coincide, with a two second uncertainty; and hence that a reasonable range for the clock offset ϕ_c is the interval (5.7, 9.7) s. But they might have been equally justified in supposing that the *centers* of the two bursts should coincide, with an uncertainty of, perhaps, three seconds. This would have given RTM another 6-s window to search for coincidences.

(b) *Length of the data set.* One of the most puzzling aspects of the RTM analysis of the KII data is that they do not stay with the 2-hour data set, but again give themselves freedom to choose its length: the one hour of data from 2 h to 3 h UT. There is no *a priori* reason for this, (all the particle data for the days 22 and 23 February 1987 were supplied by the Kamiokande group [9]) and no explanation is given, so we can only conclude that they have done it because it gives better results in the KII analysis. Indeed, in the one analysis where RTM repeat their analysis for the full 2 h, the search for triple coincidences, their probability goes up by one or two orders of magnitude when they revert to the period they analysed in Ref. [8]. We can therefore expect that similar corrections of 1 or 2 orders of magnitude will apply to all probabilities quoted in the paper.

(c) *Change of analysis methods.* RTM use analysis methods for the Kamiokande–gravitational wave coincidences that are *different* from those that they used for the Mt. Blanc–gravitational wave coincidences. In fact they use several analysis methods. They do include the same net-excitation test, but for reasons that they do not explain, they do not give a detailed study of the probability of this correlation along the lines of their Mt. Blanc study. They also do not present a threshold-coincidence analysis of the Kamiokande–gravitational-wave data along the lines of the threshold analysis of the Mt. Blanc data, i.e., by putting a threshold on the summation or product data sets and looking for coincidences with the Kamiokande particles. This would seem to have been the most natural thing to have done. Instead they present a triple threshold-coincidence analysis.

Again, in the absence of any explanation, we must assume that RTM have done this to improve the results of their analysis. We shall try to correct for this where appropriate, and where possible, we shall use RTM's original analysis methods on the data they present.

2. KII–gravitational wave coincidences independent of Mt. Blanc

(a) *Net-excitation analysis.* RTM performed the same net-excitation analysis on the Kamiokande–Rome–Maryland data as they had previously on the Mt. Blanc–Rome–Maryland data. However, RTM do not carry this through to a probability estimate. This is strange because, as we have said, such an analysis, *independent* of the Mt. Blanc data, is absolutely crucial in deciding whether this correlation effect is present in the Kamiokande independent of Mt. Blanc. Fortunately, RTM publish enough data [9] to allow us to do the appropriate analysis here.

In Fig. 1 of Ref. [9], each value of $C_+(\phi)$ is compared with only 10^3 background values, and we see that at the chosen clock offset $\phi_c = 7.7$ s there is actually one background value which is larger than C_+. By RTM's method, Eq. (4) the probability of this correlation is about 2×10^{-3}. However, as we have seen, they should not put too much weight on the probability of a single point, when neighboring points separated by 0.1 s give different probabilities. In fact *all* the other points within a window ± 0.5 s of 7.7 s have associated probabilities of 10^{-2} or more, with several at 2 or 3×10^{-2} and one at 10^{-1}. The value at exactly 7.7 s cannot be physically significant, and a probability of about 2×10^{-2} would be more representative.

RTM do correct for the number of available choices of the clock time. Because they took 41 values of the clock time in the 4-s window about their chosen delay, separated by 0.1 s, they then multiplied all their derived probabilities by 41 to correct for this. Applied to their figure of 2×10^{-3}, this gives a probability of about 0.08. We must not apply the same correction to our best estimate of the probability, since we accept only four independent values of the delay. Multiplying our first probability, 0.02, by the number of choices, four, also gives 0.08. This is a reasonable estimate of the significance of the Mt. Blanc-type correlations that exist in the 1-h set of KII data selected by RTM. If in addition we allow for a further "washing out" of the correlations if we take the full 2-h set (which could give up to a factor of 10 or so as outlined above), and if we include another six independent choices of clock setting, we see that *there is no evidence whatsoever for physically significant Mt. Blanc-type correlations in the Kamiokande data.*

We would be justified in stopping here and looking no further at coincidences between the Kamiokande data and either the gravitational wave data or that from Mt. Blanc. The Mt. Blanc correlations fail the "acid test" for their physical reality by being absent from the KII data. However, for the sake of completeness we shall continue briefly to review the other RTM analyses of the KII data.

(b) *KII–Maryland–Rome triple coincidences.* Instead of doing the same sort of threshold-coincidence analysis that they did for the Mt. Blanc–gravitational-wave data, RTM instead perform a straight triple-coincidence analysis, looking for threshold crossings in both the gravitational wave data streams at times given by Kamioka particle times (with the appropriate delay). Indeed, this is the sort of analysis that we wish they had performed on the Mt. Blanc data in the first place. However, the fact that they did something different for Mt. Blanc makes it hard to offer this analysis as evidence that the Mt. Blanc correlations appear in Kamiokande.

RTM find a peak of 15 triple coincidences at a threshold of 40 K, for which they made an erroneous estimate of the probability, 8.2×10^{-6}, by using the mean rate of events appropriate to a different data set (see Ref. [9] or our Appendix). The data are also oversampled in the manner of their analysis of the Mt. Blanc–gravitational-wave data (see Appendix A 1 a). One way to attempt to correct for the oversampling is by averaging the number of triple coincidences in a window ± 0.5 s about the maximum. Using the "peak" of Fig. 3 in Ref. [9], we arrive at a "typical" 11 triple coincidences. If we use the correct expected number of 4.8 from the data set in question, then these have a Poisson probability of only 0.01 [see Eq. (7)]. If we then correct for the freedom to shift the clock correction, for the introduction of this analysis method which was not used before, and for the narrowing of the data set to 1 h (a correction of up to one order of magnitude), we see here as well that the correlations become completely insignificant. The claimed triple coincidences at other thresholds can be dealt with similarly.

3. Gravitational wave coincidences with KII and Mt. Blanc combined

In the Appendix, we criticize the fact that when the KII and Mt. Blanc data sets are combined and then analyzed in the manner in which the Mt. Blanc data set was, it is almost impossible to judge what is the independent contribution of KII to the resulting correlations. If the KII set does not exhibit the correlations independently, and we have just seen that it does not, then it cannot be expected to enhance the Mt. Blanc correlations.

We only wish to make one further remark about this: given the considerably greater sensitivity of Kamiokande to neutrinos than Mt. Blanc has, it is odd that RTM do not weight the KII neutrinos more heavily than the Mt.

Blanc neutrinos in the combined data set. If RTM were to weight the KII particles more strongly, then their lack of correlations would, of course, depress the significance of the correlations RTM find in the combined data sets.

For all these reasons, we do not feel that the analyses of the combined data sets contribute significantly to the probability estimates we have been making.

4. Conclusion: Kamiokande as a test of the Mt. Blanc correlations

We find no evidence that Kamiokande particles were correlated with the gravitational wave detectors in the way that Mt. Blanc neutrinos were. This is, as we have said, the acid test of the Mt. Blanc effect: if the correlation is not also present in the data of other neutrino detectors, then it is hard to believe that it is real. Despite the claims in the RTM papers about the existence of correlations in the KII data, the freedom they have to adjust parameters is enough to explain the weak correlations they find, on the null hypothesis that the KII data are random with respect to the gravitational wave data. *The RTM analysis of the KII data therefore provides the falsification of their hypothesis that the Mt. Blanc correlations are due to a real physical effect.*

C. Reassessment of gravitational wave–IMB correlations

The paper on the IMB correlations [11] appears to have been written as an attempt to unify the derivation of the earlier correlations as well as to find new ones in the IMB data. In fact, this paper presents the Mt. Blanc, KII, and IMB correlations as a coherent whole, adopting a single (but new) dataset length of $1\frac{1}{2}$ h for each, and using a similar delay, 1.2 s. Although these choices have been made *a posteriori*, it is still a welcome attempt to ensure comparability.

Using these new parameters, the significance assigned by RTM to the Mt. Blanc correlations is reduced by a factor of 30. The significance of the KII correlations is hardly changed for the following reason: in Ref. [9], the delay examined is 1.1 s and the clock correction for the KII detector is estimated to be +7.7 s, where the effect is strongest. However, in the new analysis [11], although they adopt a 1.2 s delay, they change the KII clock correction to +7.8 s, *thus canceling the detrimental effect of changing to the new 1.2 s delay.* RTM say the effect of this small difference in the clock correction is "negligible," and they do not even point out their previous use of 7.7 s, referring to it simply as another clock correction they could have used. However, the effect is far from negligible: we have seen, as can be read from Fig. 1 of Ref. [9], that this ad hoc 0.1 s adjustment makes a difference of a factor of about 10 in the strength of the correlation. We note also that RTM justify the new clock correction of 7.8 s because it matches the middle of the first five KII neutrino times with the middle of the first three IMB neutrino times. This vindicates our correction in Sec. V B 1 for the fact that RTM could have chosen methods of setting the clock correction other than the one they used.

Fortunately, the IMB clock is known to have been set correctly at the time of the experiment. However, in finding the correlation between the gravitational wave detectors and the IMB particle detector, RTM make three significant choices, all of which are *a posteriori* and for which no justification is offered.

Although in their earlier papers, they first emphasize that a 1.1 s delay is optimal [8], and they then refer to it exclusively as the optimal delay for the phenomenon, [9,10] the delay is changed back to 1.2 s here without explanation. RTM do not state the probability of the correlation at 1.1 s, and nor is it given in any graph. It is hard to know how these changes affect their claimed probabilities; but we have seen, e.g., in Fig. 6, that changing the delay around a correlation can make a difference of a factor of about 10 in the probability of the correlation.

We must also consider RTM's decision to change the length of the dataset again, this time to $1\frac{1}{2}$ hours, still centred on 2 h 45 m. This, they say, is based on their previous analyses of 2 h for Mt. Blanc and 1 h for Kamiokande. Certainly, this is a good compromise between the periods analyzed previously, and will not affect those earlier correlations too dramatically. But in fairness, RTM could quite easily have chosen a period of 1 h or 2 h instead, as they have used these before. In fact, they could have used any period of this order, as this is what they did when they chose to analyze 1 h of KII data instead of 2 h.

Finally, we must account for RTM's decision to apply an energy selection criterion to the particles detected during the experiment (see Appendix). They acknowledge that this choice affects the probability, but they make no attempt to justify it or to calculate the extent to which the probability is affected. Within the null hypothesis, of course, and particularly when these particles (high energy muons) could not possibly have been responsible for the effect observed in Mt. Blanc, which contained no muons [11], this choice has no basis. It is impossible for us to quantify the effect of this choice, since RTM do not give any results for the whole IMB particle population.

In summary, we must correct RTM's *a posteriori* calculation that the probability of the IMB correlations is 10^{-3} for their adjustment of the delay; for the choice of particle sample; and for the choice of the period of analysis. We think that a factor of 100 for such choices would not be unreasonable, giving a probability for the correlations of as much as $\sim 10^{-1}$.

VI. CONCLUSIONS

We have found that the Mt. Blanc neutrino data and the Rome and Maryland gravitational wave data streams show a weak correlation during the period of 2 h containing the Mt. Blanc "neutrino burst." The correlation is of such a nature that it would have been found once in similar *a posteriori* analyses of between 10 and 1000 random data sets. In addition, we believe that the particle data from Kamiokande and IMB show no compelling evidence of the same correlation with the gravitational wave data, while RTM accept that there are no similar correlations between gravitational waves and the Baksan detector. This leads us to the conclusion that the correlations found by RTM are most likely a chance fluctuation in the data.

In reaching these conclusions we have had to try to compensate for a host of choices and other biases in the original RTM analyses. These include: *a posteriori* choices of the time delay, of the threshold, and of the duration and starting time of the data set; 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; 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 the failure to apply consistently the Mt. Blanc analysis methods to data from Kamiokande and IMB. In assessing the effects of some of these choices we have been guided by our own numerical simulations of the RTM methods applied to random data sets.

The result is that we believe that the correlations, while present, are very much more likely to arise in random sets than RTM estimated. Since the Kamiokande, IMB and Baksan data do not show the same correlations, any physical model for these effects would not only need new particles and interactions; it would also have to explain how the Mt. Blanc detector could have responded while the larger Kamiokande detector did not. We feel that the correlations present in the data are sufficiently weak that they do not provide serious evidence for such new physics.

ACKNOWLEDGMENTS

We thank the following for useful conversations and correspondence: F. Dunstan, J. Hough, T. Niebauer, T. Piran, O. Saavedra, K.S. Thorne, and especially G.V. Pallottino and G. Pizzella. C.A.D acknowledges the financial support of the Science and Engineering Research Council.

APPENDIX A: REVIEW OF THE RTM ANALYSES

1. The gravitational-wave–Mt. Blanc coincidences

The first RTM analysis [7] dealt only with Rome and Mont Blanc data; the Maryland data were brought in later. On inspecting the raw gravitational wave and neutrino data near the time of the Mt. Blanc burst (see Fig. 2, which is reproduced from Fig. 2 of their paper [8]), RTM saw that a delay of 1.4 s between the two data streams would place the neutrinos in the Mt. Blanc event at times when the Rome signal is appreciably higher than the average background [7]. Although this was unexpected, the time delay was at least consistent with a massive neutrino model with an acceptably small mass for the neutrinos. The probability that the unusually high signal preceding the first neutrino should have occurred within a time interval of 3 s was 0.03, not very significant in view of the large gravitational wave energy that would be required to explain it.

However, it motivated a much more involved analysis [8] in which the two sets of gravitational wave data are collated and examined for coincidences (see Sec. A 2). The time under consideration in this analysis is the full 18 hour period, from 12 h 22 February to 6 h 23 February. During this period, the Mt. Blanc detector counted 775 events, five of which were in the neutrino burst at 2 h 52 m 37 s. The remainder represent a fairly normal background counting rate.

In searching for coincidences between the two gravitational wave data streams and a neutrino stream, one might try the fairly standard approach of applying a *threshold* criterion to the gravitational wave streams, searching for coincidences among the three streams only when *both* gravitational wave detectors are above threshold. This would treat each data stream with equal weight; and moreover, if the data were random, each stream would be expected to produce events (neutrinos or gravitational wave threshold crossings) with a Poisson distribution of arrival intervals. The statistics of such a search would be easy to analyze.

RTM do not report having performed such an analysis on the Mt. Blanc data (although they did use this method later for analyzing the Kamiokande data—see Appendix A 3). Rather, their approach is first to combine the two gravitational wave data streams into a single one by either adding or multiplying them together, and then to use two different analysis methods, one of them a threshold criterion, to compare the combined stream with the Mt. Blanc neutrino data stream. We shall describe each of the two methods and their results separately.

a. Results of the net-excitation method

RTM first examine the full 18 hours of data, calculating both $C_\times(-1.2\text{ s})$ and $C_+(-1.2\text{ s})$ for 2-h stretches of data, moved along in $\frac{1}{2}$-h steps. Each value is compared with 10^3 background values. They find the best correlation around 2 h 45 m UT on 23 February (Fig. 5, Ref. [8]), where $n = 1$ for $N_b = 10^3$. (For this value of N_b and 80 neutrinos, there are about 5000 independent data values used for generating the 1000 background values, so the background data are probably independent in this case.) Then Eq. (4) would assign this a probability of 10^{-3}.

Importantly, this figure does not change by much if the 5 ν burst is excluded from the neutrino data set: there are correlations at a 10^{-3} level with a time delay of 1.2 s even without the neutrino data that led to the suggestion of the delay. However, it is also significant that RTM use a delay of 1.2 s here rather than the 1.4 s used previously. It appears that they adjusted ϕ to get a better correlation.

To address the question of the best time delay directly, RTM next fix their attention on the 2-h window about 2 h 45 m, and for the summation statistic $C_+(\phi)$ they vary the delay time ϕ, for each value comparing it with $N = 10^5$ to 10^6 background values. The tested values of ϕ are separated by steps of only 0.1 s, far below the gravitational wave sampling time of 1 s. The delays of 1.0, 1.1, and 1.2 s all have very low probabilities from Eq. (4), smaller than 10^{-5}. Importantly, the probability at 1.4 s is only about 10^{-4}, and within ± 0.5 s of the lowest point there are probabilities as high as 10^{-2}. [We have reproduced the RTM figure in our Fig. 6(b) above, where we compared it to the results of one of our simulations.]

Given that the gravitational wave signals are sampled at 1 s intervals, differences between time delays as small as 0.1 s cannot have physical significance. The fact that the RTM probability changes by a factor larger than 10 when ϕ changes by 0.1 s gives us a measure of the confidence we can have in these probabilities. We can think of two possible causes of these large fluctuations.

First, the fluctuations may be simple small-number statistics: with typical ranking numbers of 10^5 or so, a given small change in C_* might lead to a relatively small change in n of, say, 10^3. But when we deal with points with ranking numbers near 1, a similar change in C_* would lead to a much larger *relative* change in n, say 2 or 3 or 10, leading to a much larger relative change in the inferred $p = n/N_b$. Thus, values of p substantially greater than $1/N_b$ are relatively stable against small changes in C_*, while the smallest values of p are fairly unstable. This is why we asserted in Sec. III A 1 that Eq. (4) could be unreliable if n is small.

The second possible cause of the fluctuations may be the nonindependence of the background values. We distrust values of p near or below 10^{-4}, while the most significant correlations have values of p near 10^{-6}. It may well be that for these points the correlations among the background values ensure that there are tens of background values between any two real values of C_*, leading to spuriously large variations in p from one time delay to the next.

Given these problems, it would be more prudent to infer a probability from this method (if the method is to be used at all) by taking some sort of average over delays that span a 1-s interval. If the entire interval consistently gives ranking numbers less than, say, 100 for $N_b = 10^5$, then there may well be grounds for asserting that a correlation exists that has a probability of about 10^{-3}.

However, RTM do not do this: they consistently quote the lowest probability associated with any value of ϕ, even when very nearby time delays have considerably higher values of p. This leads them to postulate probabilities as low as 10^{-6}. If a more prudent average were applied to the RTM data, the inferred probabilities would be in the range 10^{-3} to 10^{-4}.

When we discussed other problems with the RTM analysis in Sec. V, we saw that further consideration of the free choices that RTM had in their analysis (such as the delay time ϕ) increases the probability of their finding such correlations in a random data set even further.

b. Threshold-coincidence method results

By adopting a delay of 1.2 s and a threshold on the summation data stream $E_+(t)$ of 150 K, RTM find a substantial correlation. The number of gravitational wave events at this threshold is $N_{\rm GW} = 172$. As before, there are $N_\nu = 96$ neutrinos and $N_t = 7200$ sampling intervals (2 h), so the expected number of coincidences is $\bar{n} = 2.29$. The actual number observed was 13 [see Fig. 7(b)], much larger than the other values up to ± 50 s away, and with a chance probability [Eq. (7)] of about 10^{-6}.

However, this figure does not take into account the fact that RTM (as they state) searched for coincidences at a variety of thresholds, and chose 150 K only because it gave the best correlation. This clearly affects any realistic assessment of the chance probability of this correlation, and provides one of the principal motivations for our Monte Carlo study.

2. The Rome–Maryland–gravitational-wave coincidences

Although there were no improbable coincidences between the two gravitational wave detectors at the time of the Mt. Blanc event, the correlation with neutrinos over a long period of time makes it important to see if the two gravitational wave detectors were correlated with each other over this time. We give a brief summary of the RTM analysis [1].

a. Threshold-coincidence method

RTM use data covering the 36-h period from February 21, 18 h 24 m 33 s to February 23, 6 h 2 m 3 s UT 1987, which includes the Mont Blanc burst. During this period both detectors had good thermal distributions of noise. After rescaling the Maryland data by the mass ratio of the detectors (as described in Sec. III A 1), RTM set a threshold and count the number of times that both detectors are above the threshold simultaneously. The expected number can be calculated easily from the observed exponential distributions, and by calculating threshold-crossings with various delays one can test whether the data are behaving as expected.

b. Threshold-coincidence method results

RTM divide the period under consideration into two smaller periods. Period 1, of about 7 h (2.5×10^4 s) from February 22, 23 h 5 m 23 s, to February 23, 6 h 2 m 3 s, includes the Mt. Blanc burst. period 2 is an *earlier* and longer period of 10^5 s, from February 21, 18 h 24 m 33 s to February 22, 22 h 7 m 52 s, which seems to have been analyzed as a control for the analysis of period 1. RTM do not explain why they have chosen to place the division between the two periods at about 23 h 5 m 23 s.

During period 1 there was an excessive number of coincidences above a threshold of 100 K: 41 coincidences, which is 2.4 standard deviations from the mean. Although this threshold is arbitrary, RTM give an argument that the number is still large when accumulated over 26 values of the threshold from 25 K to 150 K, and that the correlation seen has a chance probability of 3.5%.

For period 2, there is no significant correlation between the two antennas: 114, only three more than would be expected (the interval is four times as long as period 1). RTM regard this as evidence that the detectors are behaving normally during period 2, and hence, it is implied, during period 1. (They do not, however, assess the probability that one would come so close to the expected value for this period.)

Given that the gravitational wave detectors show an unusually high rate of coincidence for the period that includes the Mt. Blanc correlations described earlier, it is important to ask whether the gravitational wave correlations can by themselves account for the gravitational wave–neutrino threshold coincidences during the same period. In other words, if neutrinos arrive randomly but the gravitational wave detectors are correlated (for whatever reason, even by chance), do we expect the number of neutrino–gravitational wave coincidences that are seen?

The answer is clearly no. The number of coincidences between the gravitational wave detectors above 80 K (giving a summation energy of 160 K) is less than 20 (two standard deviations) more than would be expected by chance in the 7-h period 1. During the 2 h of the neutrino–gravitational wave coincidence analysis, this would probably give only five gravitational wave events, and the chances of their being in coincidence with a random neutrino is very small. They could explain less than 0.1 of the observed 13 neutrino–gravitational wave coincidences at a summation threshold of 150 K. It seems, therefore, that the neutrino-gravitational wave threshold coincidences occur primarily when one gravitational wave detector is well below the excitation level of the other, and are probably not associated with the coincidences tested here.

It is not clear whether the net excitation correlation of gravitational waves and Mt. Blanc neutrinos is affected by the gravitational wave–gravitational wave correlation. It is possible that the small excess of gravitational wave coincidences at most thresholds and at zero delay will raise the value of $C(\phi)$ compared to the background, because the "signal" gravitational wave values are taken at the same time, whereas the background values are taken at different times. This may marginally increase $C(\phi)$ and hence decrease n.

3. The gravitational wave–KII–Mt. Blanc coincidences

a. Net-excitation analysis of Kamiokande data

RTM first apply the net-excitation analysis method to the Kamiokande–gravitational-wave coincidence problem. For both the summation and product statistics, the best correlation occurs at a time delay of $\phi = 6.6$ s, which they take to be composed of a clock-offset adjustment $\phi_c = 7.7$ s and an intrinsic time-delay of -1.1 s, consistent with the Mt. Blanc–gravitational- wave time delay. They give a rough estimate of the chance probability of this correlation of about 10^{-3}.

b. Net-excitation analysis of merged particle data sets

RTM next do something new. They merge the set of 105 Kamiokande particles in this 1 h stretch with the set of 48 Mt. Blanc neutrinos during the same period, to give 153 particles in all. They apply the net-excitation method to this set, obtaining $C_+(\phi)$ see [Eq. (1)], with results similar to those for the Mt. Blanc data alone. Using Eq. (2) to generate a background, they get $n = 3$ for $N_b = 2 \times 10^6$, so their probability estimate for this correlation would be about 1.5×10^{-6}. Of course, the Mont Blanc neutrinos will have contributed to this, so it cannot be an independent statistic.

RTM attempt to remove this dependence on the Mt. Blanc analysis by shifting the KII signals by large random times, thereby giving a control set, where the KII signals are certainly not expected to be correlated with the Mt. Blanc neutrinos or the gravitational wave signals. They recalculate $C_+(\phi)$ and n for $N_b = 2 \times 10^6$, and obtain $n = 23,942$, a higher value than that of the "correlated" data set, indicating a weaker correlation, which can only have come about from the Mt. Blanc neutrinos. They take the ratio of these n values, viz. $3/23,942 = 1.25 \times 10^{-4}$, to be the experimental probability that the Kamiokande data's contribution to the correlation is purely chance. They then correct this for the fact that they have chosen the best value of ϕ_c from the 41 values they would have allowed themselves (steps of 0.1 s in the 4-s window for ϕ_c) by multiplying this probability by 41, arriving at a value of 5×10^{-3}, similar to the previous net-excitation probability.

It is not clear why taking the ratio of these two numbers should produce the probability of Kamiokande's "additional contribution," if any, to the Mt. Blanc–gravitational wave coincidences, and RTM make no attempt to prove this. In any case, the probability they arrive at is not very significant; but even here we must raise the same doubts as before that using steps of 0.1 s for time delays is unphysical when the data are sampled at 1 s intervals. RTM attempt to compensate for this correction by allowing for 41 independent choices for the time delay, but it is not clear to us that this compensation is correct. We have argued above that they should instead average the probability values within a 1-s window, and this could give a much larger correction, since their method starts from the unphysically small value of C_+ at the best time delay.

c. Triple-coincidence analysis

RTM examine thresholds of 30, 40, 50, and 60 K for gravitational wave events, on each detector. They do not explain why they choose these thresholds, which are much smaller than half of the threshold of 150 K that they adopted for the summation statistic in the Mt. Blanc analysis. Indeed, the triple coincidences become much less significant for the higher thresholds.

The expected number of triple coincidences, given a uniform distribution of arrival times, is

$$\bar{n} = \frac{N_\nu N_M N_R}{T^2} \tag{A1}$$

similar to Eq. (6), with N_ν the number of particle events; N_M and N_R the number of Maryland and Rome gravitational wave events, respectively; and T the length, in seconds, of the data set (here 3600). For the threshold of 40 K, this works out to be 4.8; for a threshold of 60 K it is 0.97. The actual analysis gives, for $\phi_c = 7.7$ s, 15 coincidences at 40 K and 5 at 60 K. The "raw" probabilities of these are, respectively, 1.5×10^{-4} and 3.2×10^{-3}, using Eq. (7) (RTM quote only the former, the smaller of the two). RTM would compensate for the freedom to choose a time shift by multiplying each by 41, thereby producing numbers comparable with the earlier tests of the Kamiokande data. RTM go further than this. They calculate a Poisson probability for the number of observed coincidences, but for the expected number $\bar{n}$ they use the expected number taken from an *earlier* period of time. RTM seem to think that it is better to take the mean from the earlier data, since there is no suggestion that it contains real events, and they heavily emphasize the resulting smaller probability. But this is clearly not right: the Poisson probability formula [Eq. (7)] only gives the probability of a given number of coincidences in a data set as against the expected number of events *in the data under discussion*, and using a mean from another data set will give an incorrect result. In fact, during the hour under consideration the number of Kamiokande particle events was some 25% higher than the hourly particle rate during the comparison earlier period. Therefore, RTM's contention that for the earlier comparison period, "the statistical properties of the data are very similar to those of the period of analysis" [9], is, in this sense at least, seriously incorrect. Furthermore, for the case of the 40 K threshold, using the wrong expected number (3.7) leads to a probability estimate of 8.2×10^{-6}, as compared to 1.5×10^{-4} for the actual mean number (4.8).

RTM thus quote probabilities about a factor of 20 smaller than the correct ones. However, they clearly have some doubts about this calculation.

d. Merged triple coincidence analysis

Finally, choosing the offset $\phi_c = 7.7$ s, RTM again put the Mt. Blanc and Kamiokande particles together and search for triple coincidences between all the particles and the two gravitational wave detectors. When setting a threshold of 40 K, and comparing the numbers of triple coincidences in the 2 h to 3 h UT window of 23 February with the respective average numbers of triple coincidences taken from the period 12 h to 24 h of 22 February, RTM estimate the Poisson probability of the triple coincidences found to be 2.7×10^{-8}, again using a mean taken from an earlier data set. At other choices of threshold they estimate the triple-coincidence probability to be between around 10^{-5} to 10^{-7}. This is again clearly dependent on the earlier Mt. Blanc–gravitational wave coincidences, though no attempt is made this time to correct for this. It also suffers from the fact that the number of Kamiokande particles is higher in the hour under analysis than in the comparison hours earlier. We do not see any way of using this analysis to estimate the independent contribution of the KII particles to the probability of the correlation.

e. Summary

Displaying some caution regarding these analyses, RTM prefer to adopt their earlier value of around a few times 10^{-4} [9] as their estimate of the significance of the independent support that the Kamiokande data give to the coincidences already found between the Mt. Blanc detector and the gravitational wave antennas. We have reassessed this claim in Sec. V B.

f. Note added after submission

After submission of this paper, it was pointed out to us that Ref. [10] does give a clue as to what would happen if one performed a threshold-coincidence analysis of the Kamiokande–gravitational wave data, for the original period used, *viz.* 1 h 45 m to 3 h 45 m UT. From Table II of Ref. [10], there appear to be only four coincidences of Kamiokande particles with gravitational wave data above a summation threshold of 150 K. This compares with an expected number of 5.0, obtained from: Eq. (6), from doubling $N_\nu = 105$ from the 1-h period given in Table III of Ref. [10], and from $N_{\rm GW} = 172$, given in Eq.(12) of Ref. [8]. If this were true, there would certainly be no correlation of the Mt. Blanc threshold coincidence type in the Kamiokande data, at above the chance level.

This is only what one can infer from the data presented in Refs. [9,10], which are slightly (but for these purposes, not seriously) incomplete. It would be helpful for RTM to publish the actual threshold coincidence analysis of the Kamiokande–gravitational wave data if they have not done so already.

4. The gravitational wave–IMB–Mt. Blanc coincidences

On receiving the IMB particle data, RTM analyzed them together with the gravitational wave data and the Mt. Blanc and Kamiokande particle data [11]. For brevity, we shall concentrate here on the simple coincidences between the IMB particle signals and the gravitational wave data, since only these could provide an independent confirmation of the apparent Mt. Blanc–gravitational wave correlations.

a. Net-excitation analysis of IMB data

RTM apply the same analysis technique, the net-excitation method, to the IMB signals. This time they choose to analyze only the $1\frac{1}{2}$-h period of data from 2 h 0 m to 3 h 30 m of February 23. RTM choose here to apply an energy selection criterion to the particle data. They define a quantity E(IMB), the visible energy per event given by multiplying the number of photoelectrons detected during the event by 1 MeV; 1 MeV being, very roughly, the energy deposited in the detector by one photoelectron [11]. RTM then choose to analyze only those signals whose visible energies fall in the range

$$3\text{ GeV} \leq E(\text{IMB}) \leq 6\text{ GeV} \tag{A2}$$

based on their inspection of Fig. 12 (which is Fig. 5 in Ref. [11]). The peak in Fig. 12 is due to single muons that cross the entire apparatus. This choice is very curious for the following reasons.

(1) RTM have not previously made such a selection of particle detector data based on particle energies, although for the Kamiokande data RTM adopted the same threshold that is used by the Kamiokande group in their own analysis [11].

(2) The choice seems to suggest that the muons are responsible for the correlation, while RTM have not yet offered a consistent particle-based model for any of the correlations; the choice is therefore completely *ad hoc.*

(3) The principle that new data from IMB should be subjected to the same analysis as that which found correlations in the Mt. Blanc data is somewhat compromised.

(4) If the particles in Fig. 12 really do comprise two separate populations, and if one were particularly interested, *a priori*, in the peak population, the natural energy interval to examine would be that where *the "peak" population departs from the underlying curve*, i.e., between 3 GeV and about 8.5 GeV. RTM set their cutoffs where the peak departs from the underlying curve at the lower limit (3 GeV) and where the overall number drops back down to the level at the 3 GeV limit (6 GeV). RTM thus exclude those particles between 6 GeV and 8.5 GeV, even though they belong to the same population which RTM claim they are selecting. This criterion has no scientific basis, even were it correct to separate the two populations.

(5)The Mt. Blanc data, where the effect was first seen, and where the effect is still claimed to be strongest, contain no muons [11].

For the selected particles, and adopting a delay 1.2 s, RTM find a correlation whose probability of 9×10^{-4} is assessed in the usual way for the net-excitation method. The figure for the correlation at a delay of 1.1 s is not given. For an *advance* of 1.8 s, RTM find a correlation of claimed probability of 10^{-3}. We have reassessed this analysis in Sec. V C.

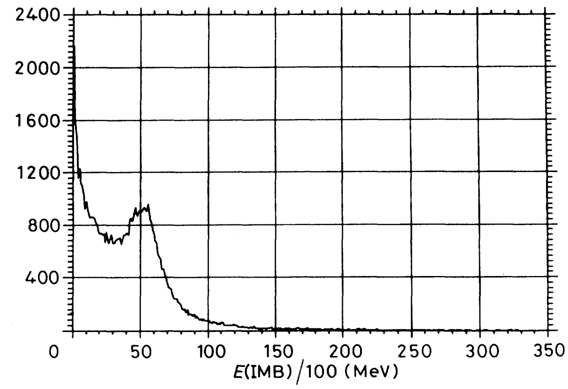

FIG. 12. The energy distribution of the IMB particles during the period in question. RTM choose to consider only those between 3 GeV and 6 GeV. (Reproduced from Ref. [11] with permission.)

[1] E. Amaldi, P. Bonifazi, S. Frasca, M. Gabellieri, D. Gretz, G. V. Pallottino, G. Pizzella, J. Weber, and G. Wilmot, in *Fourth George Mason University Workshop in Astrophysics: SN1987A in the LMC*, edited by M. Kafatos and A. Michalitsianos (Cambridge University Press, New York, 1988).

[2] K. Hirata, T. Kajita, M. Koshiba, M. Nakahata, Y. Oyama, N. Sato, A. Suzuki, M. Takita, Y. Totsuka, T. Kifune, T. Suda, K. Takahashi, T. Tanimori, K. Miyano, M. Yamada, E. W. Beier, L. R. Feldscher, S. B. Kim, A. K. Mann, F. M. Newcomer, R. Van Berg, W. Zhang, and B. G. Cortez, Phys. Rev. Lett. **58**, 1490 (1987).

[3] R. M. Bionta, G. Blewitt, C. B. Bratton, D. Casper, A. Ciocio, R. Claus, B. Cortez, M. Crouch, S. T. Dye, S. Errede, G. W. Foster, W. Gajewski, K. S. Ganezer, M. Goldhaber, T. J. Haines, T. W. Jones, D. Kielczewska, W. R. Kropp, J. G. Learned, J. M. LoSecco, J. Mattews, R. Miller, M. S. Mudan, H. S. Park, L. R. Price, F. Reines, J. Shultz, S. Seidel, E. Shumard, D. Sinclair, H. W. Sobel, J. L. Stone, L. R. Sulak, R. Svoboda, G. Thornton, J. C. van der Velde, and C. Wuest, Phys. Rev. Lett. **58**, 1494 (1987).

[4] E. N. Alexeyev, L. N. Alexeyeva, I. V. Krivosheina, and V. I. Volchenko, Phys. Lett. B **205**, 209 (1988).

[5] M. Aglietta, G. Badino, G. Bologna, C. Castagnoli, A. Castellina, V. L. Dadykin, W. Fulgione, P. Galeotti, F. F. Kalchukov, B. Kortchaguin, P. V. Kortchaguin, A. S. Malguin, V. G. Ryassny, O. G. Ryazhskaya, O. Saavedra, V. P. Talochkin, G. Trinchero, S. Vernetto, G. T. Zatsepin, and V. F. Yakushev, Europhys. Lett. **3**, 1315 (1987).

[6] 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, Refs. [8,10] some authors from Moscow have contributed. See the individual citations for the full author lists.

[7] E. Amaldi, P. Bonifazi, M. G. Castellano, E. Coccia, C. Cosmelli, S. Frasca, M. Gabellieri, I. Modena, G. V. Pallottino, G. Pizzella, P. Rapagnani, F. Ricci, and G. Vannaroni, Europhys. Lett. **3**, 1325 (1987).

[8] M. Aglietta, G. Badino, G. Bologna, C. Castagnoli, A. Castellina, W. Fulgione, P. Galeotti, O. Saavedra, G. Trinchero, S. Vernetto, E. Amaldi, C. Cosmelli, S. Frasca, G. V. Pallottino, G. Pizzella, P. Rapagnani, F. Ricci, M Bassan, E. Coccia, I. Modena, P. Bonifazi, M. G. Castellano, V. L. Dadykin, A. S. Malguin, V. G. Ryassny, O. G. Ryazhskaya, V. F. Yakushev, G. T. Zatsepin, D Gretz, J. Weber, and G. Wilmot, Nuovo Cimento C **12**, 75 (1989).

[9] E. Amaldi, M. Bassan, P. Bonifazi, M. G. Castellano, E. Coccia, C. Cosmelli, S. Frasca, D. Gretz, I. Modena, G. V. Pallottino, G. Pizzella, P. Rapagnani, F. Ricci, J. Weber, and G. Wilmot, in *Proceedings of the Fourteenth Texas Symposium on Relativistic Astrophysics*, Dallas, Texas, 1988, edited by E. J. Fenyves [Ann. N. Y. Acad. Sci. **571**, 561 (1989)].

[10] M. Aglietta, A. Castellina, W. Fulgione, G. Trinchero, S. Vernetto, C. Castagnoli, P. Galeotti, O. Saavedra, E. Amaldi, S. Frasca, G. V. Pallottino, G. Pizzella, P. Rapagnani, F. Ricci, P. Astone, C. Cosmelli, M. Bassan, E. Coccia, I. Modena, P. Bonifazi, M. G. Castellano, M. Visco, G. Badino, G. Bologna, V. L. Dadykin, F. F. Khalchukov, I. V. Korolkova, P. V. Kortchaguin, V. A. Kudryatzev, A. S. Malguin, V. G. Ryassny, O. G. Ryazhskaya, V. F. Yakushev, G. T. Zatsepin, D. Gretz, J. Weber, and G. Wilmot, Nuovo Cimento C **14**, 171 (1991).

[11] M. Aglietta, A. Castellina, W. Fulgione, G. Trinchero, S. Vernetto, P. Astone, G. Badino, G. Bologna, M. Bassan, E. Coccia, I. Modena, P. Bonifazi, M. G. Castellano, M. Visco, C. Castagnoli, P. Galeotti, O. Saavedra, C. Cosmelli, S. Frasca, G. V. Pallottino, G. Pizzella, P. Rapagnani, F. Ricci, E. Majorana, D. Gretz, J. Weber, and G. Wilmot, Nuovo Cimento B **106**, 1257 (1991).

[12] The claim that gravitational wave detectors actually have a much larger cross section than we have taken here, e.g., J. Weber, Found. Phys. **14**, 12 (1984), is wrong, as has been shown by Thorne, in *Recent Advances in General Relativity*, proceedings of a conference in honour of E. T. Newman, edited by A. Janis and J. Porter (Birkhauser, Boston, in press), and L. P. Grishchuk, Phys. Rev. D **50**, 7154 (1994); and so does not offer a way out of these problems.

[13] It would be interesting to take a more Bayesian approach to the whole analysis [see, e.g., M. G. Bulmer, *Principles of Statistics* (Dover, New York, 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.

[14] V. Trimble, Rev. Mod. Phys. **60**, 859 (1988).

[15] M. Aglietta, G. Badino, G. Bologna, C. Castagnoli, A. Castellina, V. L. Dadykin, W. Fulgione, P. Galeotti, F. F. Kalchukov, B. Kortchaguin, P. V. Kortchaguin, A. S. Malguin, V. G. Ryassny, O. G. Ryazhskaya, O. Saavedra, V. P. Talochkin, G. Trinchero, S. Vernetto, G. T. Zatsepin, and V. F. Yakushev, Europhys. Lett. **3**, 1321 (1987).

[16] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, New York, 1988).

[17] G. Pizzella and G. Pallottino (private communication).