

MONTE CARLO PREDICTION OF SPS TRIGGER RATES

Are One-Second Integrations Sufficient for Octagon Scaler Data ?

John Clem and Med Webster

Vanderbilt University

25-August-1990

Introduction

In the SPS (Short Prototype String), the PMT (Photo-Multiplier Tube) rate resulting from passing muons contributes a minute fraction of total PMT rate which is dominated by three major contributors: the bioluminescence flashes, the intrinsic PMT dark current and the K^{40} rate. Moreover, each major contributor is many orders of magnitude above the OSP (On-Ship Processor) recording limit. Hence, a condition must be imposed on the incoming data to reduce the random background without affecting the muon rate. This task is performed by the trigger processor.

The trigger processor monitors all PMT hits as they arrive aboard ship and loads each hit into its memory. The trigger processor memory locations represent a time bin and each PMT hit's time of arrival word is truncated, then used as an address to the time bin. The trigger processor continuously decodes the stored data as newly arriving data is being stored and sends output pulses when an event

satisfies a pre-set condition. Each of the 14 outputs from the trigger processor represents a different condition, and when one these conditions is satisfied, a pulse is sent out from the associated output. The pre-set conditions used during the SPS experiment either require the arrival of a designated PMT hit or require an occurrence of n -fold coincidence.

Each output from the trigger processor is counted by an associated scaler. The output from each scaler is recorded every second and is classified as the TYPE2 data. One of the trigger processor outputs is chosen as the recording device trigger to instruct the data convertor to record the selected data. These data become what is classified as the TYPE1 data. The TYPE1 data contains the PMT information that occurred within the time period centered about a trigger pulse. This data is used to reconstruct the muon's trajectory. The TYPE2 data contains the statistics of PMT hit occurrences which include the number of PMT single hits and 160ns (nano-seconds) n -fold coincidences integrated over a one-second period. A relationship between the PMT singles and n -folds can be established assuming the observed number of single PMT hits are uniformly distributed within a one-second integration time.

PMT Hit Generation

The observed number of PMT hits is relative to the PMT's rate through the Poisson probability distribution. Therefore, the probability of a PMT producing

exactly M_i hits within a integration time period of τ can be described as

$$P(M_i) = \frac{(\lambda(t)\tau)^{M_i} e^{-\lambda(t)\tau}}{M_i!}$$

where $\lambda(t)$ is the PMT hit rate and the observe number of PMT hits M_i is a random sample from this distribution. If λ is constant within the integration time the PMT hits will be uniformly distributed in time. Assuming this is the case and the product of $\lambda\tau$ is large, we have the relationship

$$\lambda\tau = M_i \pm \sqrt{M_i}.$$

Furthermore, the time uniformity assumption allows us to consider another view of this system. Instead of estimating the actual PMT rates λ , we can choose to consider the number of observed PMT hits (M_i) as the number of trials of detecting PMT hits within a shorter time period ($dt < \tau$) less than the integration time τ . The latter view will be considered in this paper.

Theory

The n-fold coincidence dependence on the single PMT hits may be clearer through a different representation. Consider the analogy of representing a PMT hit by a ball and representing the PMT identity by the color of the ball. The time bins of the trigger processor are represented by cells which can contain a ball of any color. Using this representation, the solution to the below problem

can be applied to the expected coincidence rate from a random background in the trigger processor.

The Problem

If a population of M balls of indistinguishable size are divided into 7 subsets of distinguishable color ($M=m_1+m_2+m_3+m_4+m_5+m_6+m_7$) and then randomly distributed about N cells, what is the expected number of cells to have k or greater (where $3 \leq k \leq 7$) balls of distinguishable color ?

Solution 1¹

Assuming the size of each subset population m_i is independent of the other subsets, each probability distribution of color i deposited in a specific cell can be calculated separately from the other populations. Each ball can be dealt with two different ways, either it is deposited in a specific cell or it is not. Since the fraction of cells outside a specific cell is $\frac{N-1}{N}$, the probability distribution of the number of balls in a specific cell, which follows Binomial statistics, is

$$P_i(j) = C_j^{m_i} \left(\frac{1}{N} \right)^j \left(\frac{N-1}{N} \right)^{(m_i-j)}$$

where $C_j^{m_i} = \frac{m_i!}{j!(m_i-j)!}$ and j is number of balls in a specified cell. Therefore the probability a specific cell is completely depleted of color i is

$$P_i(j=0) = ((N-1)/N)^{m_i}$$

and

$$P_i(j > 0) = 1 - ((N - 1)/N)^{m_i}$$

is the probability of an occupied cell where j is the number of balls of the same color. The conditional probability of requiring a ($Nfold$) multiple colored occurrence in a specified cell can be constructed using Multinomial statistics. Since the probability distribution for each color can be different we must sum over all color combinations.

$$P(Nfold = 3) = \sum_{h=1}^5 \sum_{i=h+1}^6 \sum_{k=i+1}^7 P_h(j > 0) P_i(j > 0) P_k(j > 0) \prod_{m(\neq h \neq i \neq k)=1}^7 P_m(j = 0)$$

$$P(Nfold = 4) = \sum_{h=1}^5 \sum_{i=h+1}^6 \sum_{k=i+1}^7 P_h(j = 0) P_i(j = 0) P_k(j = 0) \prod_{m(\neq h \neq i \neq k)=1}^7 P_m(j > 0)$$

$$P(Nfold = 5) = \sum_{h=1}^6 \sum_{i=h+1}^7 P_h(j = 0) P_i(j = 0) \prod_{m(\neq h \neq i)=1}^7 P_m(j > 0)$$

$$P(Nfold = 6) = \sum_{h=1}^7 P_h(j = 0) \prod_{m(\neq h)=1}^7 P_m(j > 0)$$

$$P(Nfold = 7) = \prod_{m=1}^7 P_m(j > 0)$$

The expect number of cells to have a prescribed event is the number of cells

N multiplied by the probability of the event.

$$\left\langle N(\geq N_{fold}) \right\rangle = N \sum_{K=N_{fold}}^7 P(K)$$

Solution 2

Consider the case where the population is without color and each ball is placed uniformly among N cells one at a time. Using Binomial statistics, as discussed in the Solution 1, the distribution of the number of balls in a cell is

$$P_{NC}(J) = C_J^M \left(\frac{1}{N} \right)^J \left(\frac{N-1}{N} \right)^{(M-J)}$$

where J is total number of balls in a specified cell.

Out of this sample of J balls, we now consider the case where the balls have distinguishable color. To determine the probability of observing an exact number of different colors within a group of J balls we must construct all the possible arrangements which satisfy this condition and then divide by the total number of unconditional arrangements.

The 1-fold case occurs when all J balls have the same color. The total number of single color arrangements is

$$\sum_{i=1}^7 C_J^{m_i}$$

and has a probability of

$$P_1^J = \sum_{i=1}^7 \Gamma(1)$$

where

$$\Gamma(1) = \frac{C_J^{m_i}}{C_J^M}.$$

In the higher order cases, the calculations begin to get increasingly complicated. Not only do we sum over all possible color combinations, but also all possible population combinations for each specific color combination.

The 2-fold case occurs when all J balls have exactly 2-different colors. Since many arrangements are possible, it is easier to view the system when we investigate a specific case. If we consider the example where J=9, we can then write all population combinations for a given color combination as shown below,

$$\begin{pmatrix} a & a & a & a & a & a & a & a \\ a & a & a & a & a & a & a & i \\ a & a & a & a & a & a & i & i \\ a & a & a & a & a & i & i & i \\ a & a & a & a & i & i & i & i \\ a & a & a & i & i & i & i & i \\ a & a & i & i & i & i & i & i \\ a & i & i & i & i & i & i & i \\ i & i & i & i & i & i & i & i \end{pmatrix}$$

where a and i represent a different color and each column represents a different population combination. Therefore, the total sum of all arrangements is written

as

$$\sum_{L=1}^{J-1} \sum_{a=i+1}^7 \sum_{i=1}^6 C_L^{m_i} C_{J-L}^{m_a}$$

and has a probability of

$$P_2^J = \sum_{L=1}^{J-1} \sum_{a=i+1}^7 \sum_{i=1}^6 \Gamma(2)$$

where

$$\Gamma(2) = \frac{C_L^{m_i} C_{J-L}^{m_a}}{C_J^M}.$$

The 3-fold case occurs when all J balls have exactly 3-different colors. By induction from 2-fold case result, the occurrence of the 3-fold case has a probability of

$$P_3^J = \sum_{N=1}^{J-2} \sum_{L=N+1}^{J-1} \sum_{p=a+1}^7 \sum_{a=i+1}^6 \sum_{i=1}^5 \Gamma(3)$$

where

$$\Gamma(3) = \frac{C_N^{m_i} C_{L-N}^{m_a} C_{J-L}^{m_p}}{C_J^M}.$$

The 4-fold case has a probability of

$$P_4^J = \sum_{S=1}^{J-3} \sum_{N=S+1}^{J-2} \sum_{L=N+1}^{J-1} \sum_{q=p+1}^7 \sum_{p=a+1}^6 \sum_{a=i+1}^5 \sum_{i=1}^4 \Gamma(4)$$

where

$$\Gamma(4) = \frac{C_S^{m_i} C_{N-S}^{m_a} C_{L-N}^{m_p} C_{J-L}^{m_q}}{C_J^M}.$$

The 5-fold case has a probability of

$$P_5^J = \sum_{T=1}^{J-4} \sum_{S=T+1}^{J-3} \sum_{N=S+1}^{J-2} \sum_{L=N+1}^{J-1} \sum_{b=q+1}^7 \sum_{q=p+1}^6 \sum_{p=a+1}^5 \sum_{a=i+1}^4 \sum_{i=1}^3 \Gamma(5)$$

where

$$\Gamma(5) = \frac{C_T^{m_i} C_{S-T}^{m_a} C_{N-S}^{m_p} C_{L-N}^{m_q} C_{J-L}^{m_b}}{C_J^M}.$$

The 6-fold case has a probability of

$$P_6^J = \sum_{U=1}^{J-5} \sum_{T=U+1}^{J-4} \sum_{S=T+1}^{J-3} \sum_{N=S+1}^{J-2} \sum_{L=N+1}^{J-1} \sum_{c=b+1}^7 \sum_{b=q+1}^6 \sum_{q=p+1}^5 \sum_{p=a+1}^4 \sum_{a=i+1}^3 \sum_{i=1}^2 \Gamma(6)$$

where

$$\Gamma(6) = \frac{C_U^{m_i} C_{T-U}^{m_a} C_{S-T}^{m_p} C_{N-S}^{m_q} C_{L-N}^{m_b} C_{J-L}^{m_c}}{C_J^M}.$$

And finally the 7-fold case, which is when all 7-different colors are present among the J balls, has a probability of

$$P_7^J = \sum_{V=1}^{J-6} \sum_{U=V+1}^{J-5} \sum_{T=U+1}^{J-4} \sum_{S=T+1}^{J-3} \sum_{N=S+1}^{J-2} \sum_{L=N+1}^{J-1} \Gamma(7)$$

where

$$\Gamma(7) = \frac{C_V^{m_1} C_{U-V}^{m_2} C_{T-U}^{m_3} C_{S-T}^{m_4} C_{N-S}^{m_5} C_{L-N}^{m_6} C_{J-L}^{m_7}}{C_J^M}.$$

We have determined the probability distribution of finding J balls in a given cell which will be represented as $P_{NC}(J)$ and then the probability distribution of finding K different colors from a group of J balls which will be represented as P_K^J . Hence probability of finding K different colors in a cell is

$$P(K = 4) = \sum_{i=4}^M P_{NC}(i) P_4^i$$

$$P(K = 5) = \sum_{i=5}^M P_{NC}(i) P_5^i$$

$$P(K = 6) = \sum_{i=6}^M P_{NC}(i) P_6^i$$

$$P(K = 7) = \sum_{i=7}^M P_{NC}(i) P_7^i.$$

And again the expect number of cells to have a prescribed event is the number of cells N multiplied by the probability of the event.

$$\left\langle N(\geq N_{fold}) \right\rangle = N \sum_{K=N_{fold}}^7 P(K)$$

We have two solutions which represent the relationship between the PMT singles and n -fold coincidences. Since an analytical comparison of the two solutions is nontrivial to calculate, each solution is compared through analyzing Monte Carlo data.

Comparison of Solutions using Monte Carlo Data

The two solutions were incorporated into two separate fortran subroutines which were both called from a common main program. The main program generated random PMTs counts and passed the same values into each subroutine.

The subroutines returned their calculated expected number coincidences and were compared to each other. The resulting data is presented in figures 1,2,3 and 4. The two different solutions are consistent for n-folds ranging from 3-fold to 6-fold. This consistency strengthens our confidence of having an accurate theory that predicts the expected number of n-fold coincidences.

Trigger Processor Simulation

PMT Single Random Hits

The Trigger Processor simulator begins by generating seven values which represent the number of observe hits that occurred within the first second from each of the seven PMTs. The number of observed PMT hits is generated from fitted curves of the observe single rate distribution measured during the SPS experiment. The PMT hits are uniformly distributed among 6.25×10^6 bins each representing a 160ns window. After the distribution is completed, each individual bin is checked as the counters representing a prescribed condition are updated. After scanning all time bins, the counters values and the PMT single hits values are written into a data file. The counters are then cleared and the process is repeated for the next second.

Muon Events

The probability of an atmospheric muon event within a one-second period at a

four kilometer ocean depth is also included among the randomly distributed PMT single hits. The probability of an event occurring follows Poisson statistics and the average muon rate is determined from the expected flux angular distribution integrated over a 30m radius sphere.

The muon trajectories were generated in the same direction as the axis of a cylinder and uniformly within the cross-sectional area. The radius of the cylinder is 30m, which is large enough to include the 3-fold events with the standard attenuation length curve². The cylinder could be rotated about an axis perpendicular to the cylinder's axis with the pivot point position at OM4 location, which is the center of SPS.

The trajectory calculation gives the mean of the number of photo-electrons (PE) expected at each PMT. The Poisson distribution is used to generate the number of PEs for the event. If the random Poisson generator produces a zero, the detector did not observe the Čerenkov light from the passing muon. Moreover, a 20 percent³ hit loss was also included to simulate the observed data-loss of the SPS. However, if a hit survives both tests, an observed pulse width is generated for the detector. A mean pulse width from the PMT is determined from the mean number of PEs using the relationship obtained from the PMT calibrations. The calibrations show that the pulse width distribution is Gaussian and the width is dependent on the mean. So the pulse width is varied according to calibrations to generate a sample which is the observed pulse width.

The observed pulse width is used to slew the mean time of arrival value. The PMT time resolution calibrations show the time of arrival distribution is Gaussian and the width is a function of the mean number of PEs, therefore a Gaussian smearing is also performed on the time of arrival. The delay times of the underwater fiber optic cables from each OM to the string bottom controller (SBC) are then added to the associated observed time of arrival.

A random time offset is added to PMT hits to determine the exact time within the one-second period the muon event occurred. The resulting PMT time of arrival data are then truncated to established an associated time bin among the 6.25×10^6 bins.

Comparison of the N-Fold Theoretical Prediction with Monte Carlo Data and SPS Data

The calculation assumes the n-fold coincidences are random accidentals, while the muon events must be accounted as separate constant source. Hence an offset must be included to account for this effect. Since the probability of a 3-fold or greater coincidence occurring in a given time bin is small and the number of time bins are very large, the probability distribution of the n-folds in a one-second integration period follows Poisson statistics. Therefore,

$$P(K_i) = \frac{\left((R_i + b)T\right)^{K_i} e^{-(R_i + b)T}}{K_i!}$$

where R_i is the predicted number of n-fold hits per second, K_i is the observed number of n-folds, b is an offset to account for the muon events and T is set at one second observations.

The Likelihood Function

$$\mathcal{L} = \prod_{i=1}^J P(K_i),$$

where J is the number of one-second interval observations, is the product of the probability for each observation. Each component represents the probability that the observed number of n-fold hits (K_i) is a member of a Poisson distribution whose mean is set by the expected number of n-folds $((R_i + b)T)$. The maximum of the Likelihood Function in terms of b is determine by minimizing the below function

$$\ell = -\log \mathcal{L}.$$

The resulting minimization process provides the most probable value for (b_m) the muon rate. Ignoring the higher order terms in the Likelihood Function near its maximum, the curvature is related to error in (b_m) the muon rate as shown in the below equation.

$$\frac{2}{db_m^2} \approx d^2(\log \prod_{i=1}^J P(b = b_m)) / db^2.$$

These approaches were applied in fitting the SPS data at 4km depth and three

different Monte Carlo data sets. Each Monte Carlo data set was generated under different conditions.

Monte Carlo Uniform PMT Rates without Muon Events

The first Monte Carlo data set is limited to only uniform random background hits without muon events. If the prediction theory of n -fold coincidences is correct, the resulting b fits to this data set should provide values which are consistent with zero regardless of the n -fold requirement. Figure 5 shows the b fit results and reconfirms the prediction theory. The errors on the 3-fold are very large since many 3-fold events are produced even during the minimum PMT single rates resulting in a large lever-arm near the origin. On the other hand, very few 6-folds occurred producing a small sample of observed successes.

Monte Carlo Uniform PMT Rates with Muon Events

The second Monte Carlo data set is uniform background hits with muon events. The resulting b values to this data set are proportional to the SPS muon effective area for each n -fold trigger requirement. Figure 6 indicates the effective area decreases with increasing n -folds triggers. This effect is due to a decreasing likelihood of detecting distance muons near the one PE level and an increasing likelihood of losing a hit in the SPS data loss simulation with a higher n -fold trigger. Moreover, the chance of a n -fold event straddling two or more time bins increases with a higher n -fold trigger. The error bars indicate the signal-to-noise

ratio decreases with decreasing n-folds. It also should be noted that the b value for the 5-folds is consistent with the 4km 5-fold muon rates⁴ from the SPS fitted trajectories with a 85% fitting efficiency correction.

SPS data at 4km

The b fits to the SPS scalar data set is shown in figure 7. The resulting b values are much larger than one would expect from results shown in figure 6. The 6-fold results are consistent, but the differences increase with decreasing n-folds. One possibility is that the SPS PMT rates are not uniform within a one-second observation. This explanation motivated the need to generate Monte Carlo data with a non-uniform background to determine its effects on our predictions.

Bioluminescence Spikes

The third Monte Carlo data set is uniform background noise with muon hits and one millisecond noise spikes. The short length of the spikes is unrealistic for most bioluminescence, but will suffice as a comparison to a uniform distribution. The origin of the spikes is the result of light sources in the vicinity of the SPS. The locations of the sources are uniformly distributed within a 5m radius of the SPS and the distribution density falls off as one over the radius squared from 5m to 25m. The size of the spike is dependent on the location of the light sources to account for the one over the radius squared effect, the attenuation effect and angular sensitivity of the SPS PMT's. An upper limit of the spike size is included

to avoid spikes which exceed a factor 25 of the uniform background rate.

As shown in figure 8, this b fit curve for different n-folds is very similar to SPS data in figure 7. The differences between these two sets of data is probably due to the narrow noise spikes and the uncertainty in the bioluminescence population density distribution used in the Monte Carlo data. However, it demonstrates how noise spikes can hinder the extraction of the muon rates from the scaler data. Since the length of the noise spikes are less than integration time of the scaler, this information is lost and can not be accounted for.

Conclusion

Our predictions of the SPS accidental trigger rates are correct as long as the PMT hit rates are uniform within the integration time. Moreover, we are able to extract the muon events from accidental trigger rates in the Monte Carlo uniform noise data and the 5-fold muon event rate is consistent with the SPS muon fitted trajectory rate. However, the SPS muon trajectory rate did not agree with its trigger scaler rate. The excess accidental trigger rates in the SPS scaler data suggest the integration time was not short enough since the PMT noise rates fluctuated within the one-second period.

Even though the SPS was ship suspended and mechanically stimulated most of the bioluminescence it observed, the bottom moored TTR4⁵ also observed occasional large noise spikes. Therefore, these large noise spikes should be ob-

served by the Octagon array. Since a comparison check of the muon rate from fitted trajectories will be needed, our recommendation is to allow for a smaller and variable integration time in the Octagon's data acquisition, which will help provide consistent results from the scaler data.

REFERENCES

1. John Babson's Ph.D Thesis, University of Hawaii, 375
2. John Clem's Ph.D Thesis, Vanderbilt University , 88
3. John Clem's Ph.D Thesis, Vanderbilt University , 60
4. John Clem's Ph.D Thesis, Vanderbilt University , 106
5. M. Webster, *et al.*, "Mechanical Stimulation of Bioluminescence in the Deep Pacific Ocean", Submitted to **Deep-Sea Research**

FIGURE CAPTIONS

Figure 1) A comparison of the Multinomial (1) solution and the Combinatorial (2) solution of the expected number of 160ns time bins which have a 3-fold hits from seven random PMT counts from Monte Carlo data.

Figure 2) Same as last figure for the the 4-fold case.

Figure 3) Same as last figure for the the 5-fold case.

Figure 4) Same as last figure for the the 6-fold case.

Figure 5) Excess background or offset fit to a Monte Carlo data set is limited to only uniform random background hits. If the prediction theory of n-fold coincidences is correct, the resulting excess background or offset fits to this data set should provide values which are consistent with zero regardless of the n-fold requirement. This plot displays the offset fit results and therefore reconfirms the prediction theory. The errors on the 3-fold are very large since many 3-fold events are produced even during the minimum PMT single rates resulting in a large lever-arm near the origin. On the other hand, very few 6-folds occurred

producing a small sample of observed successes.

Figure 6) The second Monte Carlo data set is uniform background hits with muon events. The resulting fitted offset values to this data set are proportional to the SPS muon effective area for each n-fold trigger requirement. This figure indicates the effective area decreases with increasing n-folds triggers. This effect is due to a decreasing likelihood of detecting distance muons near the one photoelectron level and a increasing likelihood of losing a hit in the SPS data loss simulation with a higher n-fold trigger. Moreover, the chance of a n-fold event straddling two or more time bins increases with a higher n-fold trigger requirement. The error bars indicate the signal-to-noise ratio decreases with decreasing n-folds. It also should be noted that the offset value for the 5-folds is consistent with the 4km 5-fold muon rates⁴ from the SPS fitted trajectories with a 85% fitting efficiency correction.

Figure 7) The resulting offset fits to the SPS scalar data set is displayed. The resulting offset values are much larger than one would expect from results shown in figure 6. The 6-fold results are consistent, but the differences increase with decreasing n-folds. One possibility is that the SPS PMT rates are not uniform within a one-second observation.

Figure 8) The third Monte Carlo data set is uniform background noise with muon hits and one millisecond noise spikes. In the simulation, the origin of the spikes is the result of light sources in the vicinity of the SPS. As shown, this offset fit curve for different n-folds is very similar to SPS data in figure 7. This effect demonstrates how noise spikes could have caused elevated offsets in the SPS data.



