Astronomical Data Analysis Software and Systems X
ASP Conference Series, Vol. 238, 2001
F. R. Harnden Jr., F. A. Primini, and H. E. Payne, eds.

Uniform Data Sampling: Noise Reduction & Cosmic Rays

J. D. Offenberg[1], D. J. Fixsen[1], M. A. Nieto-Santisteban[2],
R. Sengupta[1], J. C. Mather[3], H. S. Stockman[2]

Abstract. As computers and scientific instruments become more complicated and more powerful (Moore's Law), we can perform astronomical observations never before contemplated. As larger data volumes are acquired, as more complex instruments are designed, and as observatories are placed in distant space locations with constrained downlink capacity, the need for automated, robust image processing tools will increase.

We present a robust, optimized algorithm to perform automated processing of array image data obtained with a non-destructive read-out. We present the derivation of the noise effects of this algorithm and compare alternative strategies.

1. Introduction

The effects of radiation and cosmic rays can be a formidable source of data loss for a space-based observatory. Several solutions to the problem of identifying and removing cosmic rays exist. We evaluate Up-the-Ramp sampling with on-the-fly cosmic ray identification and mitigation, which is described in detail by Fixsen et al. (2000) and compared to Fowler Sampling (Fowler & Gatley 1990). We concentrate on Up-the-Ramp sampling for study because it provides better signal-to-noise in what is probably the most difficult-to-measure regime, the read-noise limit. In the absence of cosmic rays, Up-the-Ramp sampling provides modestly ($\sim$ 6%) higher signal-to-noise than does Fowler Sampling (Garnett & Forrest 1993). The fact that an Up-the-Ramp sequence can be screened for cosmic rays and other glitches improves this result. Furthermore, on-the-fly cosmic ray rejection allows longer integration times which also improves the signal-to-noise in the faint limit (Offenberg et al. 2001).

The following discussion is largely an excerpt from Offenberg et al. 2001.

2. Fowler Sampling

Fowler sampling reduces the effect of read noise to $\sigma_r' = \sigma_r\sqrt{4/N}$ (for an observation sequence consisting of N samples, $N/2$ Fowler-pairs). However,

[1]Raytheon ITSS, 4500 Forbes Blvd, Lanham MD 20706

[2]Space Telescope Science Institute, 3700 San Martin Dr., Baltimore MD 21818

[3]Code 685, NASA's Goddard Space Flight Center, Greenbelt MD 20771

when a pixel is impacted by a cosmic ray during an observation, the cosmic ray essentially injects infinite variance and reduces the signal to noise to zero at that location. If we start with the Fowler sampling signal-to-noise function in the read-noise limit, from Garnett & Forrest (1993; Eqn. 6),

$$SN_F = \frac{FT}{\sqrt{2}\sigma_r}\sqrt{\frac{\eta T}{2\delta t}}\left(1 - \frac{\eta}{2}\right) = \frac{FT}{\sqrt{V_0}} \tag{1}$$

where F is the flux of the target, T is the observation time, σ_r is the read noise, η is the Fowler duty cycle, and δt is the time between sample intervals (determined by engineering or scientific constraints on the system). We note that this formula breaks down for relatively small numbers of samples (i.e., δt large with respect to T). FT is the signal, so the remaining terms are the noise, which is the square-root of the variance, V_0. If we consider two cases, "no-cosmic-ray" and "hit-by-cosmic-ray," and combine the variances according to

$$V_{comb} = \frac{V_0 P_0 W_0^2 + V_1 P_1 W_1^2}{(W_0 P_0 + W_1 P_1)^2} \tag{2}$$

we can rewrite Equation 1 as

$$SN_{FC} = \frac{FT}{\sqrt{V_{comb}}} = \frac{FT(W_0 P_0 + W_1 P_1)}{\sqrt{V_0 P_0 W_0^2 + V_1 P_1 W_1^2}} \tag{3}$$

As the weight is the inverse of the variance ($W_i = 1/V_i$), Equation 3 can be rewritten as

$$SN_{FC} = \frac{FT(\frac{P_0}{V_0} + \frac{P_1}{V_1})}{\sqrt{P_0/V_0 + P_1/V_1}} = FT\sqrt{\frac{P_0}{V_0} + \frac{P_1}{V_1}} \tag{4}$$

V_0 is the variance in the no-cosmic-ray case, taken from Equation 1, and P_0 is the probability of a pixel surviving without a cosmic ray hit. For simplicity, we define $1 - P$ to be the probability of a pixel being hit by a cosmic ray per time unit δt, so P is the probability of "survival" and $P_0 = P^{T/\delta t}$. As a cosmic ray hit injects infinite uncertainty, the variance in the cosmic ray case is $V_1 = \infty$. Plugging in to Equation 4, we get the signal-to-noise for Fowler sampling in the read-noise limited case with cosmic rays, Equation 5:

$$SN_{FC} = FT\sqrt{\frac{P_0}{V_0} + 0} = \frac{FT}{\sqrt{2}\sigma_r}\sqrt{\frac{\eta T}{2\delta t}}\left(1 - \frac{\eta}{2}\right) P^{T/(2\delta t)} \tag{5}$$

For a given integration time T and minimum read time δt, the maximum SN_{FC} occurs with duty cycle $\eta = 2/3$. If we plug this back into Equation 5, we get

$$SN_F = \frac{2}{3}\frac{FT}{\sqrt{2}\sigma_r}\sqrt{\frac{T}{3\delta t}} P^{T/(2\delta t)} \tag{6}$$

From here, it is possible to find the value of T which gives the best signal-to-noise for a single observation; it occurs at $T = -3\delta t/\ln(P)$. If, however, we consider the observation as a series of M equal observations with a specific total observation time, T_{obs}, the signal-to-noise for the series is

$$SN'_{FC} = \frac{FT\sqrt{M}}{3\sigma_r}\sqrt{\frac{2T}{3\delta t}}P^{T/(2\delta t)} = \frac{FT\sqrt{T_{obs}}}{3\sigma_r\sqrt{T}}\sqrt{\frac{2T}{3\delta t}}P^{T/2\delta t} \tag{7}$$

If we hold T_{obs} constant and find the optimum T, we find it at $-2\delta t/\ln(P)$. In either case, it is important to note that there is an optimal value for T, and extending the observation beyond that time will ruin the data.

It is worth noting that the result assumes that all cosmic ray events can be identified *a posteriori*. This is not necessarily the case, particularly when it is considered that, in the one-image case, the fraction of pixels surviving without a cosmic ray impact is $P^{-3/\ln(P)} = e^{-3} \approx 0.05$; for the multi-image case, the fraction of survivors is $P^{-2/\ln(P)} = e^{-2} \approx 0.14$. In both cases, the number of "good" pixels is so low that separating them from the impacted pixels will not be a trivial task. For example, the median operation would not be able to identify a good samples, as more than half of the samples would be impacted by cosmic rays. In practice, the detector will often saturate before this limit is reached, but this shorter integration time means that less-than-optimal signal-to-noise will be obtained.

3. Up-the-Ramp Sampling

Up-the-Ramp sampling reduces the effect of read noise to $\sigma'_r = \sigma_r\sqrt{12/N}$, for N uniformly-spaced samples with equal weighting (which is the optimal weighting for the read-noise limited case). When a pixel is impacted by a cosmic ray, the Up-the-Ramp algorithm preserves the "good" data for that pixel. The exact quality of the preserved data depends on the number of cosmic ray hits and their timing within the observation. For example, a cosmic ray hit which just trims off the last sample in the sequence has minimal impact compared to a cosmic ray hit that occurs in the middle of the observation sequence. The variance of a Uniformly-sampled sequence with N_i samples is proportional to $1/N_i(N_i+1)(N_i-1)$. If an Up-the-Ramp sequence is broken into i chunks by a cosmic ray, the variance becomes

$$V_i = V_U \frac{N(N+1)(N-1)}{\sum_{j=0}^{i}(N_j)(N_j+1)(N_j-1)} \tag{8}$$

When there are zero cosmic ray events, of course, $V_0 = V_U$. If there is one cosmic ray event during the sequence, the variance becomes

$$V_1 = V_U \frac{N(N+1)(N-1)}{N_i(N_i+1)(N_i-1)+(N-N_i)(N-N_i+1)(N-N_i-1)} \tag{9}$$

If we assume (as is reasonable) that the cosmic ray events are randomly distributed over time and find the expectation value for all values of $0...N_i...N$, we find that the typical $V_1 \approx V_U * 2$ (plus a small term in N^{-1}, which we will ignore for simplicity). If we perform a similar computation for two cosmic ray events, we find that $V_2 \approx V_U * 10/3$ (again, plus lower-order terms which we ignore). In general, we find that it is possible to find a valid result with a finite variance for any sequence broken up by cosmic ray events provided we have at least two consecutive "good" samples (for all practical purposes, we can ignore the situation where this is not the case). To simplify the following, we consider only three cases: The no-cosmic-ray case $V_0 = V_U$, the one-cosmic-ray case $V_1 = 2 * V_U$ and all multiple-cosmic-ray cases combined as one, $V_{2+} = V_U/\epsilon^2$, where ϵ^2 is a small but non-zero number, roughly 0.3.

The Up-the-Ramp signal-to-noise function for the read-noise limited case (Garnett & Forrest 1993; Eqn. 20) is

$$SN_U = \frac{FT}{\sqrt{2}\sigma_r}\sqrt{\frac{N^2-1}{6N}} = \frac{FT}{\sqrt{V_U}} \tag{10}$$

We combine the variances in the three possible cases with the three-case equivalent to Equation 2, and thus arrive at

$$SN_{UC} = FT\left(\frac{P_0}{V_0} + \frac{P_1}{V_1} + \frac{P_{2+}}{V_{2+}}\right)^{1/2} = \frac{FT}{\sqrt{V_U}}\left(P_0 + \frac{P_1}{2} + \epsilon P_{2+}\right)^{1/2} \tag{11}$$

where P_i is the probability of a pixel being impacted by i cosmic rays during the integration. We note, as did Garnett & Forrest, that there would be no reason to limit the number of samples to anything less than the maximum possible number, so we can set $N = T/\delta t$. Using the definition of P described earlier, $P_0 = P^{T/\delta t}$, $P_1 = (T/\delta t)(1-P)P^{T/\delta t-1}$ and $P_{2+} = 1-(P_0+P_1)$. Putting these values back into Equation 11, we get:

$$SN_{UC} = \frac{FT}{\sqrt{2}\sigma_r}\sqrt{\frac{T^2-\delta t^2}{6T\delta t}}\left[(1-\epsilon)P^{T/\delta t} + (1-\epsilon)\frac{T}{\delta t}(1-P)P^{T/\delta t-1} + \epsilon\right]^{1/2} \tag{12}$$

If we seek the maximum value of SN_{UC} with respect to T, we find that SN_{UC} is strictly increasing if $T \geq \delta t$ (otherwise we would have an integration shorter than one sample time, which would be useless), $P \neq 0$ and $0 < \epsilon < 1$ (both of which are true by construction). This result applies whether we are considering one independent integration or a series of observations to be combined later. As the derivative is strictly positive, the signal-to-noise continues to increase with the sample time, although as $T \rightarrow \infty$, the gain in signal-to-noise asymptotically approaches zero. So, extending the observing time while using Up-the-Ramp sampling with cosmic ray rejection does not damage the data (although we might be spending time with little or no gain). As noted earlier for the Fowler-sampling case, there is an optimal observing time, beyond which further observation reduces the overall signal-to-noise.

References

Fixsen, D. J., et al. 2000, PASP, 112, 1350

Fowler, A. M. & Gatley, I. 1990, ApJ, 353, L33

Garnett, J. D. & Forrest, W. J. 1993, Proc. SPIE, 1946, 395

Offenberg, J. D., et al. 2001, PASP, 113, in press