MIT Kavli Institute

Chandra X-Ray Center

MEMORANDUM
August 15, 2013

To: Jonathan McDowell, SDS Group Leader

From: John C. Houck, SDS

Subject: A better row-loss upper limit for `destreak`

Revision: 2.0

URL: `http://space.mit.edu/cxc/docs/docs.html#s4streak_loss`

File: destreak_limit_criterion.tex

1 Overview

Images of the `ACIS-8` chip show a variable pattern of linear streaks that fall along rows of pixels with constant `CHIPY`. The `destreak` tool identifies and optionally removes these streaks by exploiting the fact that each streak occurs within one frame and deposits multiple events along one `CHIPY` row of a node. For low count-rate observations, source events rarely cluster in this way. The `destreak` filtering algorithm[1] and the properties of the streak events[2] are described elsewhere.

Recently, a small number of observations (*e.g.* obsid 15543) were affected by such a high streak rate that pipeline processing with `destreak` failed to remove all of the streaks. In these cases, re-running `destreak` with a larger value of `max_rowloss_fraction` filtered the streaks much more completely without significant loss of source events, suggesting that the default value of `max_rowloss_fraction` might be too conservative.

Motivated by this experience, the main purpose of this memo is to consider how `destreak` might be made more effective. Because the effect of the `max_rowloss_fraction` parameter is not as straightforward as one might naively assume, §2 attempts to clarify precisely what this parameter does. Based on processing the available public data using a range of `max_rowloss_fraction` values, it appears that one simple approach is to increase `max_rowloss_fraction` from 5×10^{-5} (the current default) to `max_rowloss_fraction` $= 2 \times 10^{-4}$. Section §3 provides some justification for this suggestion. In §4, I describe a different limiting criterion based on a different parameter that may provide more sensitive control over the effect of `destreak`.

[1] `http://space.mit.edu/cxc/docs/docs.html#s4streak_alg`
[2] `http://space.mit.edu/cxc/docs/docs.html#s4streak_prop`

2 Effect of the max_rowloss_fraction parameter

By definition, a streak is the occurrence of *more than* $N_{\rm streak}$ events in a single CHIPY row, of a single node, within a single frame. While the value of $N_{\rm streak}$ can be specified by the user via the max parameter, it is more common to let destreak choose the value of $N_{\rm streak}$ as follows.

On the first pass through the input event file, $N_{\rm streak} = 1$, meaning that the occurrence of two or more events in a single row/node/frame is interpreted as a streak. Therefore, on the first iteration destreak identifies as many streaks as possible. A streak is removed by discarding all events in the relevant row/node/frame.

If the fraction of rows discarded in the first pass exceeds max_rowloss_fraction on a given node, then the tool iteratively increases $N_{\rm streak}$ on that node and re-applies the new threshold until the fraction of rows discarded on that node falls below max_rowloss_fraction.

If the fraction of rows discarded in the first pass does not exceed max_rowloss_fraction, then the tool attempts to derive an optimal value of $N_{\rm streak}$ by exploiting the fact that the number of events per streak on a given node is an exponentially distributed random variable. Using the population of streaks identified on the first pass, destreak fits the observed streak distribution with an exponential function and then computes a new threshold value, $N_{\rm streak,opt}$, that will optimally distinguish between the streak events and the non-streak background. destreak then makes a second pass through the data, re-identifying streaks using the optimal threshold value $N_{\rm streak}$ =$N_{\rm streak,opt}$.

An alternative implementation might, in the absence of a user-specified value of the max parameter, first derive $N_{\rm streak,opt}$ and only then impose the max_rowloss_fraction limit. Given that the $N_{\rm streak,opt}$ approach usually works very well, as can be verified by running destreak with max_rowloss_fraction $= 1$, this alternative implementation might be a useful improvement.

In adjusting the max_rowloss_fraction parameter within the current implementation, it may be helpful to keep in mind the following characteristics. First, the max_rowloss_fraction parameter changes the behavior of destreak only when the first pass through the data identifies enough streak events to exceed the max_rowloss_fraction threshold. Second, because the effect of the destreak filter is directly dependent on the discrete, nonlinear parameter $N_{\rm streak}$, the max_rowloss_fraction parameter does not give fine control over the action of the filter. Third, imposing an overly restrictive upper limit on the value of max_rowloss_fraction can reduce the effectiveness of the streak filter, preventing it from identifying streaks with small numbers of events per streak, (e.g. $\sim$3-5 events per streak).

3 Analysis of Existing Data

To assess the impact of an increase in max_rowloss_fraction, I used destreak to filter all public, non-calibration, TE mode ACIS observations with ACIS-8 turned on, a set of 4758 obsids (3786 imaging, 972 grating) as of this writing in August 2013. I also examined all public, non-calibration, CC mode ACIS observations with ACIS-8 turned on, a set of 513 obsids (275 imaging, 238 grating). The streak rate distribution seen in the CC mode data is consistent with that seen in the TE mode data, but because the TE mode data is so much more common, all results presented here refer to the TE mode data unless otherwise stated.

For consistency with current pipeline usage of destreak, the status bits in the input event file were cleared and destreak was run with no mask (mask=) so that all events in the input event file were considered as potential streak events. destreak was run on all of the input data using max_rowloss_fraction $= 5 \times 10^{-5}$ (the current default), 10^{-4}, 3×10^{-4}, and 10^{-3}.

For each `max_rowloss_fraction` value, I examined the effective exposure time loss per `CHIPY` row associated with the action of the `destreak` filter. The `timefile` output from `destreak` records the total exposure time discarded from each row on each node. For each row, the exposure time loss reflects the number of frames in which a streak was flagged in that row. To establish notation, consider an obsid with `EXPTIME`$=\tau$ and `EXPOSURE`$= T = N\tau$. Suppose that in m separate frames, `destreak` flags a streak in row $y =$ `CHIPY` of node i. Discarding these m rows reduces the total exposure in that row by an amount $\Delta T = m\tau$, corresponding to the loss of a fraction, $f_y = \Delta T/T = m/N$, of the total exposure in row y.

After examining all the publicly available grating and imaging data including both TE-mode and CC-mode, it became apparent that more of the TE-mode grating observations were affected by an increase in `max_rowloss_fraction` than were the other types of observations. For this reason, I use the TE mode grating data to illustrate the dependence of exposure time loss on `max_rowloss_fraction` (see Figure 1). The effect of a particular value of `max_rowloss_fraction` on all types of data examined is shown in Figure 2.

For the majority of TE mode grating observations, setting `max_rowloss_fraction` $\leq 10^{-4}$ keeps the maximum exposure time loss $\lesssim 1\%$ (see Figure 1a and Figure 1b), while setting `max_rowloss_fraction` $\gtrsim 3\times 10^{-4}$ can sometimes lead to a loss of $\gtrsim 1\%$ of the exposure time in one or more `CHIPY` rows (see Figure 1c and Figure 1d). Inspection of Figures 2a -2d shows that with `max_rowloss_fraction` $= 3\times 10^{-4}$, most imaging and grating observations, whether TE-mode or CC-mode, lose $\lesssim 1\%$ of exposure in any `CHIPY` row. One exception is obsid 1078 (TE mode, imaging) – an unusual case in which an extremely bright source, LMC X-1, was placed on node 0 of `ACIS-8` for a short (739 sec) calibration observation. In this case, the maximum exposure loss on node 0 is 273.6 sec, or 37% of the total exposure time. However, for this obsid, this problem persists even with `max_rowloss_fraction` $= 5\times 10^{-5}$.

Based on these results, it appears that simply increasing the default value of the limit parameter to `max_rowloss_fraction` $= 2\times 10^{-4}$ will ensure more effective and consistent streak filtering while also limiting any potential loss of source events to a level comparable to or smaller than the calibration uncertainty in the effective area. The same parameter default should work for both timed-exposure (`TE`) and continuous clocking (`CC`) mode, and for both grating (HETG or LETG) and non-grating data.

4 A Better Limit Criterion

In the longer term, a better solution may be to use a more sensitive limit criterion. A problem with the current implementation is that the limit criterion on the total number of rows discarded is effectively node-averaged, making it less sensitive for distinguishing the presence of a bright source from an unusually high streak rate. A more sensitive criterion would impose a limit based on the maximum exposure time lost in any single row or, equivalently, the maximum number of frames in which any single row is discarded.

To define the limit criterion more precisely, consider an observation in which N frames are read out, with M rows per frame so that a total of NM rows are read out. Define $\delta_{iyn} = 1$ if row y is discarded from node i in frame n, and $\delta_{iyn} = 0$, otherwise. The total number of streak rows discarded from node i in the entire observation is then,

$$\delta_i = \sum_{y=1}^{M}\sum_{n=1}^{N}\delta_{iyn}, \tag{1}$$

so that a fraction, $\delta_i/(NM)$, of all rows read out are discarded from node i. This fraction may be interpreted as the mean row-loss fraction, or row-loss probability, for node i, averaged over the duration of the observation and over all CHIPY rows. To simplify notation, drop the explicit node index and define the mean row-loss fraction, $\mathrm{mean}(f_y) \equiv \delta_i/(NM)$. The parameter max_rowloss_fraction specifies the maximum allowable value of $\mathrm{mean}(f_y)$; the current default is max_rowloss_fraction $= 5 \times 10^{-5}$.

Using the notation established above, the number of frames in which row y is discarded from node i is

$$\delta_{iy} = \sum_{n=1}^{N} \delta_{iyn}, \tag{2}$$

so that the maximum number of frames discarded from any row on node i may be written as $\mathrm{max}_y(\delta_{iy})$. Expressing this as a fraction of the total number frames, and dropping the explicit node index, we can define the maximum single-row loss fraction as, $\mathrm{max}(f_y) \equiv \mathrm{max}_y(\delta_{iy})/N$. After multiplying the numerator and denominator by the frame time, τ, this fraction may also be interpreted as the maximum fraction of the total exposure time lost by any single row.

As long as the fraction of the total exposure time lost in any single row is small compared to the calibration uncertainty of the relevant effective area, any loss of source events must be negligible — with the exception of sources that are detected only through a small number of short, bright flares. However, events from such faint, flaring dominated sources are already at risk of removal by the hot-pixel/afterglow filter and in any case, a search for such sources using data strongly affected by ACIS-8 streaks is questionable at best.

For the vast majority of practical cases, a conservative limit on the maximum exposure loss in any single row should be sufficient to guarantee that destreak has done no significant harm.

Implementing this new limit criterion would require only minor changes to destreak. Unfortunately, such changes would then require new regression tests, and some existing regression tests may need updating. A new destreak interface would require updates to ahelp documentation, data processing caveats and relevant threads and eventually, small changes to pipeline processing scripts may be required.

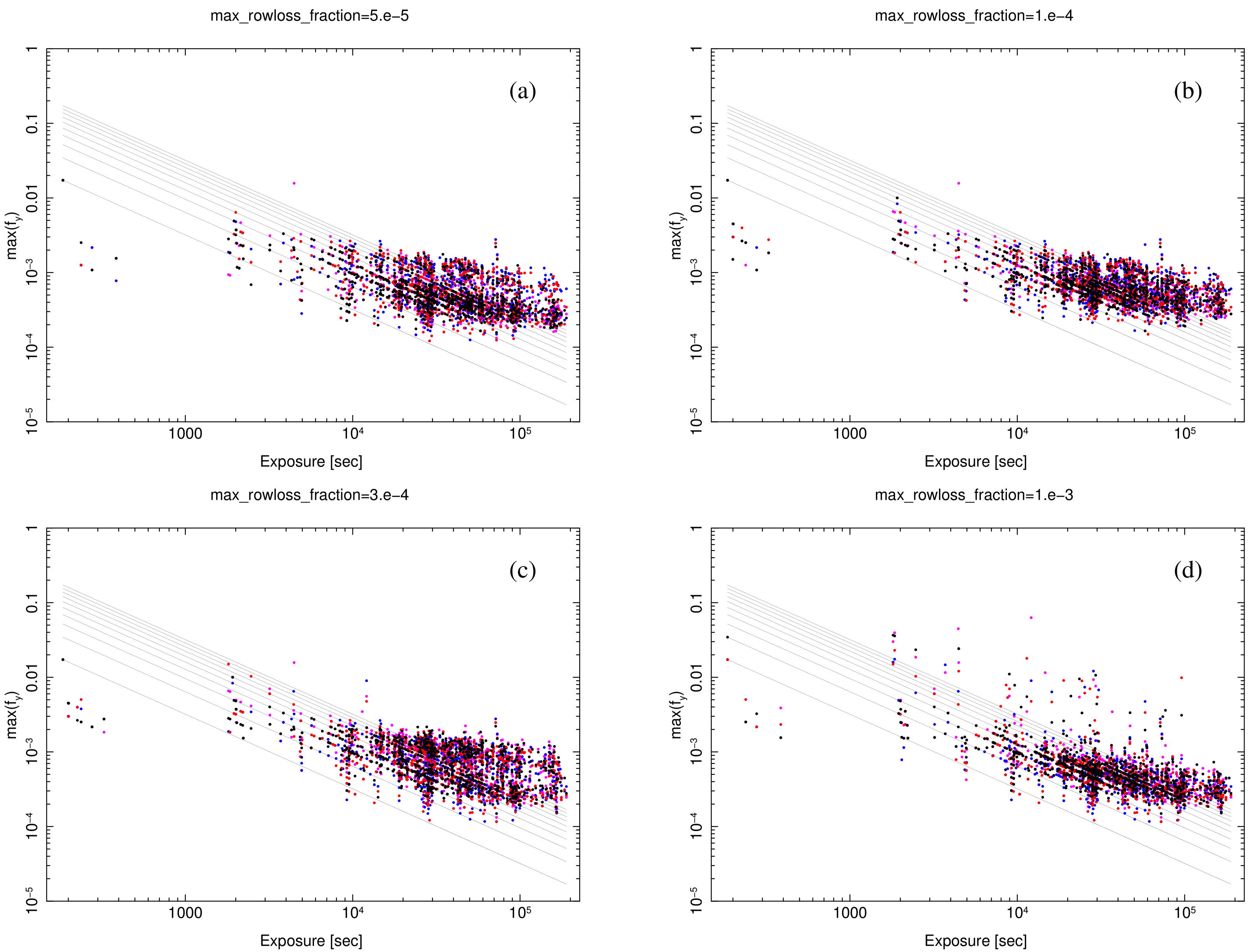

Figure 1: $\max(f_y)$ vs. exposure time for 961 TE-mode grating obsids for four different values of `max_rowloss_fraction`, (a) 5×10^{-5}, (b) 1×10^{-4}, (c) 3×10^{-4}, and (d) 1×10^{-3}. Each point corresponds to a single node from a single obsid; different colors correspond to different nodes. The grey diagonal lines correspond to a loss fraction of $3.2m/T$, where T is the `ACIS-8` exposure time in seconds and $m = 1, 2, \ldots, 10$ is the maximum number of frames discarded from a single row.

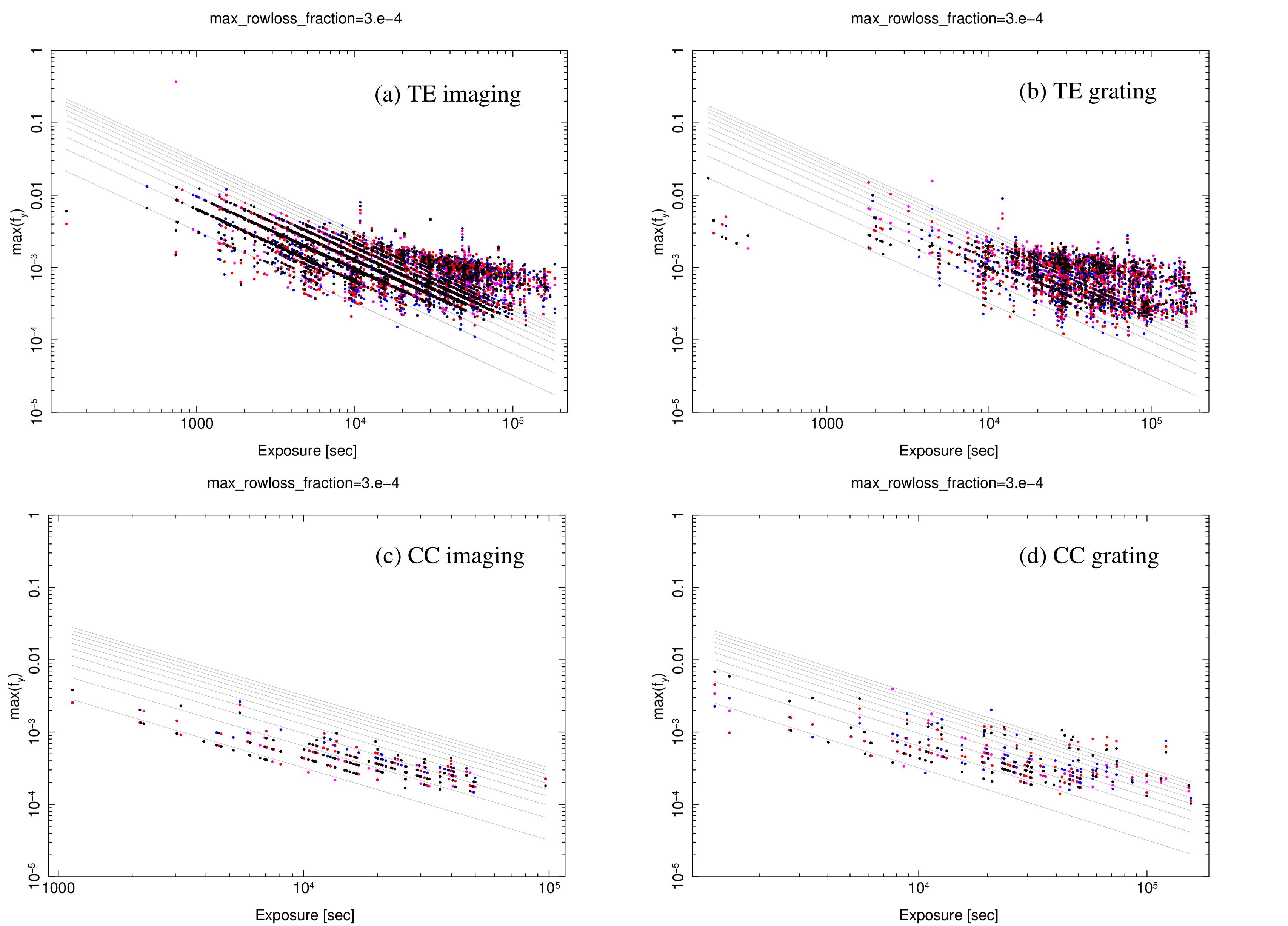

Figure 2: $\max(f_y)$ vs. exposure time using $\texttt{max\_rowloss\_fraction} = 3 \times 10^{-4}$, for TE-mode imaging data (a) and grating data (b) and for CC-mode imaging data (c) and grating data (d). Each point corresponds to a single node from a single obsid; different colors correspond to different nodes. The grey diagonal lines correspond to a loss fraction of $3.2m/T$, where T is the ACIS-8 exposure time in seconds and $m = 1, 2, \ldots, 10$ is the maximum number of frames discarded from a single row.