MIT Kavli Institute **Chandra X-Ray Center**

MEMORANDUM
April 30, 2012

To:	Jonathan McDowell, SDS Group Leader
From:	Glenn E. Allen, SDS
Subject:	`DELTOCLK spec`
Revision:	2.0
URL:	http://space.mit.edu/CXC/docs/docs.html#deltoclk
File:	/nfs/cxc/h2/gea/sds/docs/memos/deltoclk_spec_2.0.tex

Since this version of the spec differs substantially from Rev. 1.4, the changes are not highlighted in red.

1 Description

Infrequently, the values of `DELTOCLK` in exposure-records files are anomalously high or anomalously low for a particular frame of a particular node. As a result the bias- and `DELTOCLK`-adjusted pulse-height data is inaccurate. Recall that

$$\texttt{PHAS}[i,j]_k = \texttt{RAW\_PHAS}[i,j]_k - \texttt{BIAS}[i,j] - \texttt{DELTOCLK}_{k-1}, \tag{1}$$

where $\texttt{PHAS}[i,j]_k$ is the adjusted pulse height for pixel $[i,j]$ of an event island in frame k, $\texttt{RAW\_PHAS}[i,j]_k$ is the unadjusted pulse height, $\texttt{BIAS}[i,j]$ is the bias value of the pixel, and $\texttt{DELTOCLK}_{k-1}$ is the value of `DELTOCLK` for the corresponding node in the preceding frame.

This memo describes an algorithm to identify anomalous values of `DELTOCLK`. The identification process accommodates the following features. The initial values of `DELTOCLK` are typical invalid (i.e. 4095). There can be steep positive or negative gradient in the values of `DELTOCLK` at the beginning of an observation until the temperature of a detector stabilizes. Small scale, periodic variations occur in some of the data. One feature that is not handled by the algorithm, at least using the default parameter settings, is the case when there is a large number of frames between consecutive values of `DELTOCLK` ("frame gaps"). An examination of the data suggests that such cases are rare and that the sparsely sampled values of `DELTOCLK` are not anomalous. The well-sampled data immediately preceding and following a gap is handled well.

Level 0 data files remain unchanged (i.e. include the anomalies), but Level 1 files have pulse-height and overclock data that are corrected, if possible, to compensate for the anomalies. Events that are adversely affected by anomalies have a `STATUS` bit set to one and are excluded from Level 2 event files. A list of the anomalies is included in exposure-statistics files so that users can compensate for any changes to the exposure time.

2 Input

1. Level 0 pulse-height event islands

2. Level 0 DELTOCLKs

3. Level 0 EXPNOs

3 Output

1. Level 1 pulse-height event islands that have been updated to compensate for anomalies

2. Level 1 event STATUS bit information that has been updated to indicate whether or not events are affected by anomalies

3. A Level 1 exposure-statistics file with OVRCLOCK values that have been updated to compensate for anomalies

4. A Level 1 exposure-statistics file that includes a list of the anomalies in a BADOCLK HDU

4 Parameters

1. `infile,f,a,"",,,"Name(s) of input event-data file(s)"`

2. `outfile,f,a"",,,"Name of output event-data file"`

3. `exrfile,f,a,"",,,"Name(s) of input exposure-records file(s)"`

4. `expstatfile,f,a,"",,,"Name of output exposure-statistics file"`

5. `numframes,i,h,5,3,1001,"Nominal number of data points in the sliding time window used to smooth the data (must be odd and no less than minnumframes)"`

6. `maxframegap,i,h,11*,1,1001,"Maximum number of frames between consecutive data points in window"`

7. `minnumframes,i,h,3,3,1001,"Minimum number of data points in window"`

8. `deltoclkthresh,i,h,3†,1,4096,"Minimum offset in adu that is considered anomalous"`

9. `numiter,i,h,3,1,10,"Number of iterations performed to smooth the data"`

5 Processing

1. The validity of the input is verified. The `infile` and `exrfile`(s) must exist. The data in the `exrfile`(s) must be valid. A valid `exrfile` is one where each row contains a numerical value for EXPNO and four numerical values for DELTOCLK (i.e. one value for each node). The values of EXPNO must increase from one row to the next. The values of the parameters `numframes`, `maxframegap`, `minnumframes`, `deltoclkthresh`, and `numiter` must be in their valid ranges. The parameter `numframes` must be an odd number that is greater than or equal to `minnumframes`. If one or more of these conditions is not satisfied, then a warning message is produced and the data are not searched for anomalous values of DELTOCLK.

*The largest DTYCYCLE used to date is 10. Furthermore, as shown in Figure 2, the frame numbers in consecutive rows of data typically differ by eleven or less.

†As shown in Figures 3–5, the algorithm is relatively insensitive to the value of `deltoclkthresh` for values greater than or equal to three. Therefore, a value of three seems to represent a good balance between maximizing the number of anomalies identified and minimizing the false-positive rate.

2. To determine if one or more values of DELTOCLK is anomalous, the data in an exrfile are processed to estimate the actual values of DELTOCLK. Except as noted below, the estimates M are obtained by smoothing the data with a sliding median[‡] filter.

   (a) The computation of the value of M for the first row of data is performed last as described in item 2f.

   (b) If there are at least two rows of data, then for the second row $M_2 = \texttt{DELTOCLK}_2$.

   (c) If there are at least three rows of data and if $n \geq 2$, where

$$n = \frac{\texttt{numframes} - 1}{2}, \tag{2}$$

   then for rows k from 3 to $n{+}1$, M_k is the median of the set $[\texttt{DELTOCLK}_{k-1}, \texttt{DELTOCLK}_k, \texttt{DELTOCLK}_{k+1}]$, provided

$$\begin{aligned} \texttt{EXPNO}_k - \texttt{EXPNO}_{k-1} &\leq \texttt{maxframegap} \text{ and} & (3)\\ \texttt{EXPNO}_{k+1} - \texttt{EXPNO}_k &\leq \texttt{maxframegap}. & (4) \end{aligned}$$

   If either equation 3 or 4 is not satisfied, then $M_k = \texttt{DELTOCLK}_k$. The data at the beginning is handled differently than the rest of the data because it is not uncommon to have relatively steep DELTOCLK gradients at the beginning of an observation.

   (d) If there are at least $n + 2$ rows of data, then for rows $k \geq n + 2$, M_k is the median of the set $[\texttt{DELTOCLK}_{k-n}, \ldots, \texttt{DELTOCLK}_{k+n}]$, provided that the maxframegap constraint is satisfied. For example, if $n = 2$, then M_k is the median of $[\texttt{DELTOCLK}_{k-2}, \texttt{DELTOCLK}_{k-1}, \texttt{DELTOCLK}_k, \texttt{DELTOCLK}_{k+1}, \texttt{DELTOCLK}_{k+2}]$ provided

$$\begin{aligned} \texttt{EXPNO}_{k-1} - \texttt{EXPNO}_{k-2} &\leq \texttt{maxframegap}, & (5)\\ \texttt{EXPNO}_k - \texttt{EXPNO}_{k-1} &\leq \texttt{maxframegap}, & (6)\\ \texttt{EXPNO}_{k+1} - \texttt{EXPNO}_k &\leq \texttt{maxframegap}, \text{ and} & (7)\\ \texttt{EXPNO}_{k+2} - \texttt{EXPNO}_{k+1} &\leq \texttt{maxframegap}. & (8) \end{aligned}$$

   If equation 5 is not satisfied, then $\texttt{DELTOCLK}_{k-2}$ is excluded from the computation of the median because the frame gap is too large. If equation 6 is not satisfied, then both $\texttt{DELTOCLK}_{k-2}$ and $\texttt{DELTOCLK}_{\texttt{k}-1}$ are excluded from the computation of the median. If equation 7 is not satisfied, then both $\texttt{DELTOCLK}_{k+1}$ and $\texttt{DELTOCLK}_{k+2}$ are excluded from the computation of the median. If equation 8 is not satisfied, then $\texttt{DELTOCLK}_{k+2}$ is excluded from the computation of the median. If there are not at least minnumframes values included in the computation of the median, then $M_k = \texttt{DELTOCLK}_k$ because there is too little data to accurately estimate M_k (i.e. to determine if DELTOCLK is anomalous).

   (e) Steps 2c and 2d are repeated an additional $\texttt{numiter} - 1$ times (for a total of numiter times) to remove some of the noise from the estimates M. For these additional computations, the new values of M are computed from the previous values of M instead of from the values of DELTOCLK. That is, M_k is the median of the previous set of values $[M_{k-n}, \ldots, M_{k+n}]$ instead of the median of $[\texttt{DELTOCLK}_{k-n}, \ldots, \texttt{DELTOCLK}_{k+n}]$. This computation is subject to the same maxframegap constraints noted above. However, the constraint on the number of values required to compute the median is different. The value of M_k is only recomputed if there are numframes values instead of minnumframes values in the computation of the median. This more restrictive condition on the number of values in the computation of the median helps prevent the loss of accuracy in the value of M_k where there are more data points in the computation on one side of frame k than there are on the other side (as is the case at the beginning of a data set and at the edges of frame gaps).

[‡]Here, the median is defined as follows. If the number of elements in the set for which the median is being computed is odd, then the median is the middle value of the set after the set has been sorted. For example, the median of the set [1,1,4,5,6] is 4. If the number of elements is even, then the median is the mean of the middle two values of the sorted set. If the mean is not an integer, then the mean is truncated to obtain an integer. For example, the median of [1,1,4,5] is 2, not 2.5.

(f) If there is more than one value of DELTOCLK in the `exrfile`, then the first value $M_1 = M_2$, independent of value of DELTOCLK. If there is only one value of DELTOCLK and if $\mathtt{DELTOCLK}_1 = 4095$, then $M_1 = 0$ because such a value of 4095 is anomalous. If there is only one value of DELTOCLK and if $\mathtt{DELTOCLK}_1 \neq 4095$, then $M_1 = \mathtt{DELTOCLK}_1$.

3. Once the estimates M of the actual values of DELTOCLK are computed, the value of DELTOCLK for frame k is identified as anomalous if

$$\mathtt{DELTOCLK}_k \geq M_k + \mathtt{deltoclkthresh} \text{ or} \tag{9}$$

$$\mathtt{DELTOCLK}_k \leq M_k - \mathtt{deltoclkthresh}. \tag{10}$$

4. If the value of DELTOCLK for frame k is anomalous, then

(a) the `outfile` is modified as follows

i. The pulse-height data are computed using

$$\mathtt{PHAS}[i,j]_{k+1} = \mathtt{RAW\_PHAS}[i,j]_{k+1} - \mathtt{BIAS}[i,j] - M_k \tag{11}$$

instead of

$$\mathtt{PHAS}[i,j]_{k+1} = \mathtt{RAW\_PHAS}[i,j]_{k+1} - \mathtt{BIAS}[i,j] - \mathtt{DELTOCLK}_k \tag{12}$$

for all of the events that occur during frame $k+1$ (not frame k) on the particular CCD and node with an anomaly. Since the value of DELTOCLK is computed at the end of a frame, not the beginning, there is a one frame offset between the time when it is obtained and time when it is used. If an event occurs along the edge of a node that has an anomalous value of DELTOCLK, then only the pulse-height values $[i,j]$ of the event island that are affected by the anomaly are adjusted. The pulse heights of pixels that lie on the adjacent, good node are not adjusted. Similarly, if an event in frame $k+1$ occurs on a good node that is adjacent to a node with an anomalous value of DELTOCLK, then only the pulse heights of pixels $[i,j]$ that lie on the adjacent, bad node are adjusted. The pulse-height computation described by equation 11 is only performed if the DATAMODE is CC33_FAINT, FAINT, FAINT_BIAS, or VFAINT. The computation is not performed if the DATAMODE is CC33_GRADED or GRADED.

ii. STATUS bit 10 (of 0-31) is set to one for all of the events that are affected by an anomaly, whether the events are on the node with the anomaly or an adjacent node (see item 4(a)i).

(b) and the `expstatfile` is modified as follows.

i. The OVRCLOCKs are computed using

$$\mathtt{OVRCLOCK}_k = M_k + \mathtt{INITOCL}x \tag{13}$$

instead of

$$\mathtt{OVRCLOCK}_k = \mathtt{DELTOCLK}_k + \mathtt{INITOCL}x, \tag{14}$$

where x = A, B, C, or D for NODE_ID = 0, 1, 2, or 3, respectively.

ii. A "BADOCLK" HDU is included, which contains a list of the EXPNO, CCD_ID, and NODE_ID information for the frame immediately following frame k.

6 Caveats

1. As of the date of this memo, the tools `mkarf`, `mkgarf`, `mkinstmap`, and `mkwarf` have not been modified to include the time dependence introduced by specific nodes being bad during specific frames. Note that the bad region includes not only the node with an anomaly, but also a column in one or more adjacent nodes (see item 4(a)i).

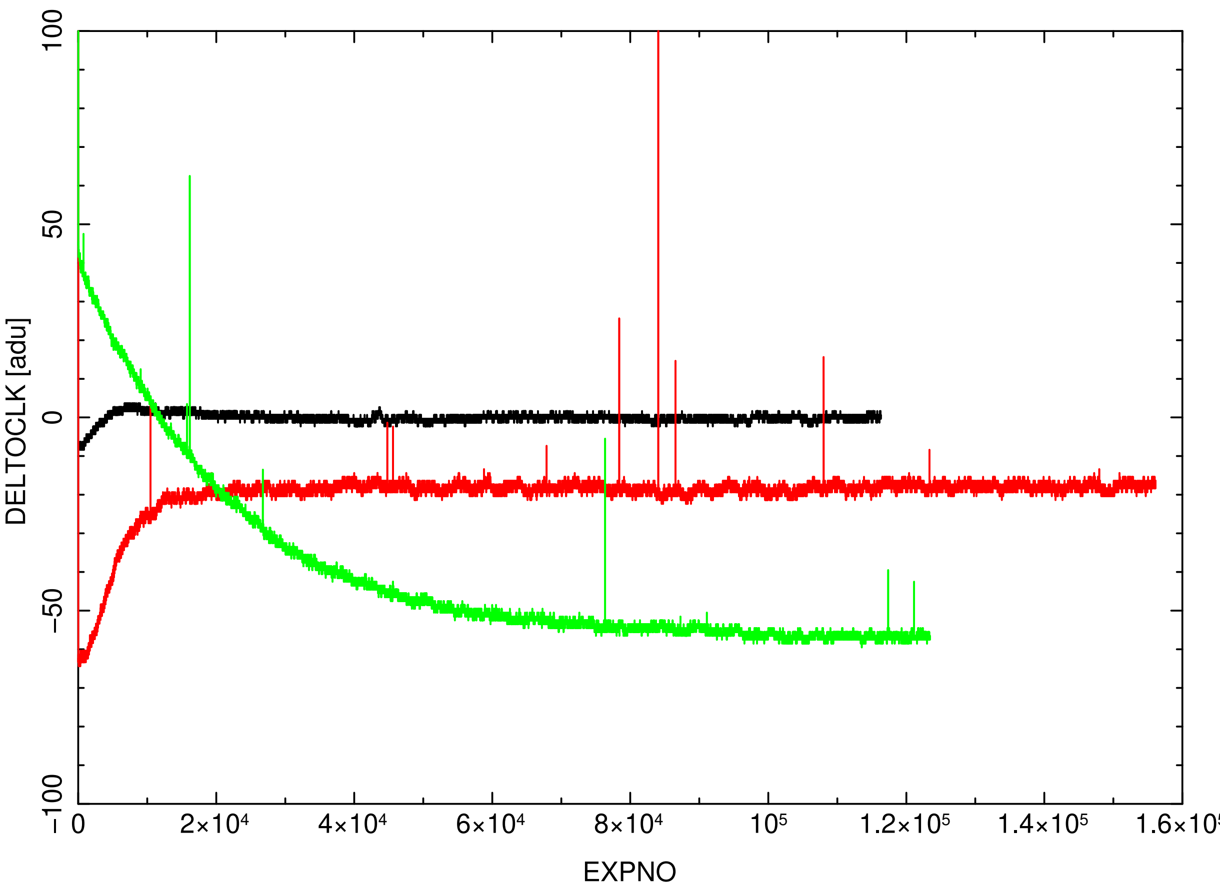

Figure 1: Three examples of the relative values of DELTOCLK as a function of frame number. The data are from exposure-statistics files and an arbitrary constant has been added to the values of DELTOCLK to improve the visual clarity. The black, red, and green curves are the data for $(\texttt{OBS\_ID}, \texttt{CCD\_ID}, \texttt{NODE\_ID}) = (1886, 7, 0)$, (11793,9,1), and (13019,9,2), respectively. Anomalous "spikes" are evident in the red and green curves. The black curve has no such anomalies.

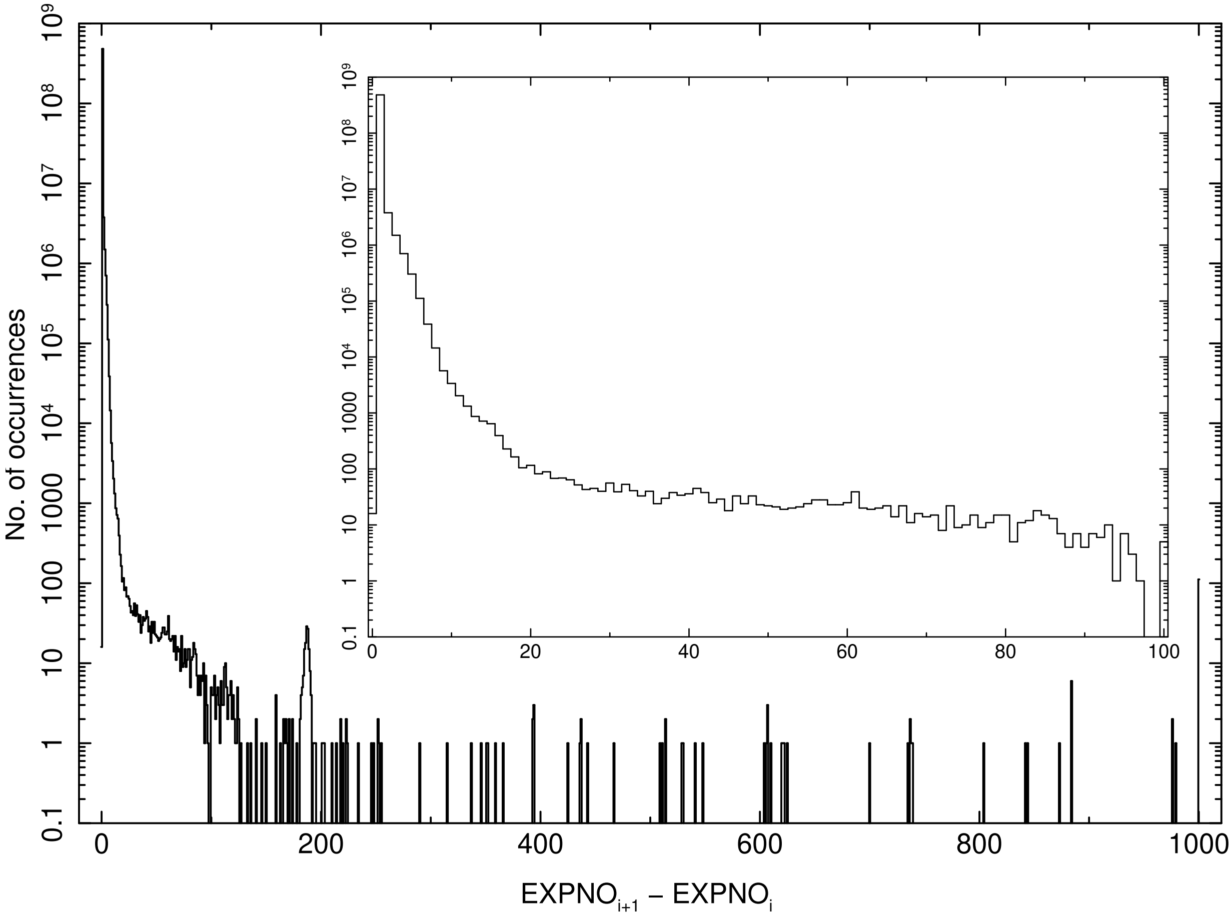

Figure 2: A histogram of the differences between consecutive frame numbers for 63,778 exposure-records files. A separate histogram was computed for each node of each file. The histogram in this figure is a sum of each of these histograms. If a difference is greater than 1000, then the difference is set to 1000. If a difference is less than 0, then the difference is set to 0. The inset histogram is a plot of the same data, except that it includes only the data in the range from 0 to 100. The histogram includes all frames.

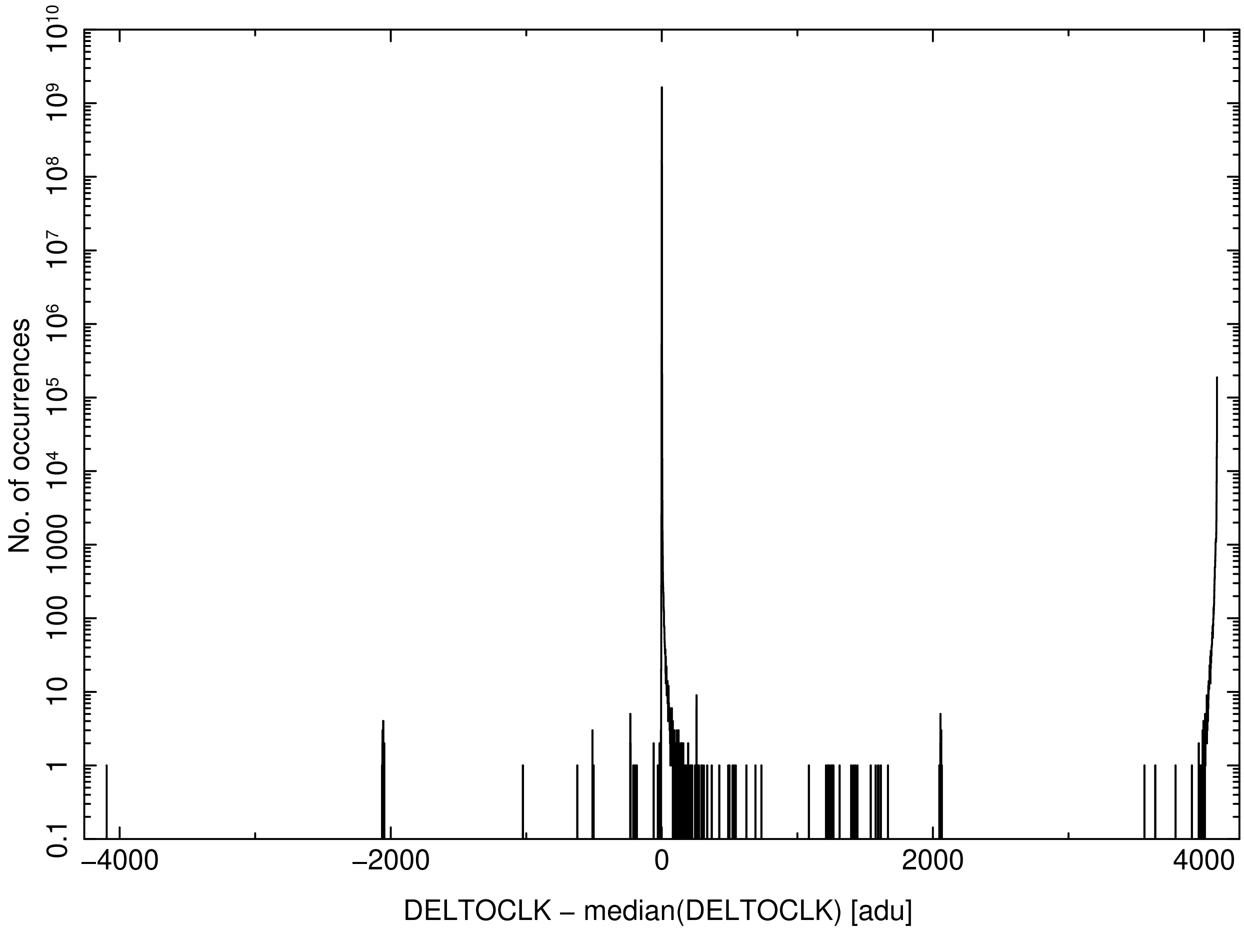

Figure 3: A histogram of the differences between the values of `DELTOCLK` and the values of the median `DELTOCLK` for 63,778 exposure-records files. A separate histogram was computed for each node of each file. The histogram in this figure is a sum of each of these histograms. The values of the median were computed using a `numframes` $= 5$, `maxframegap` $= \infty$, and `minnumframes` $= 3$. If a value of `DELTOCLK` differs by more than 4096 adu from the corresponding median, then the difference is set to 4096 or -4096 as appropriate.

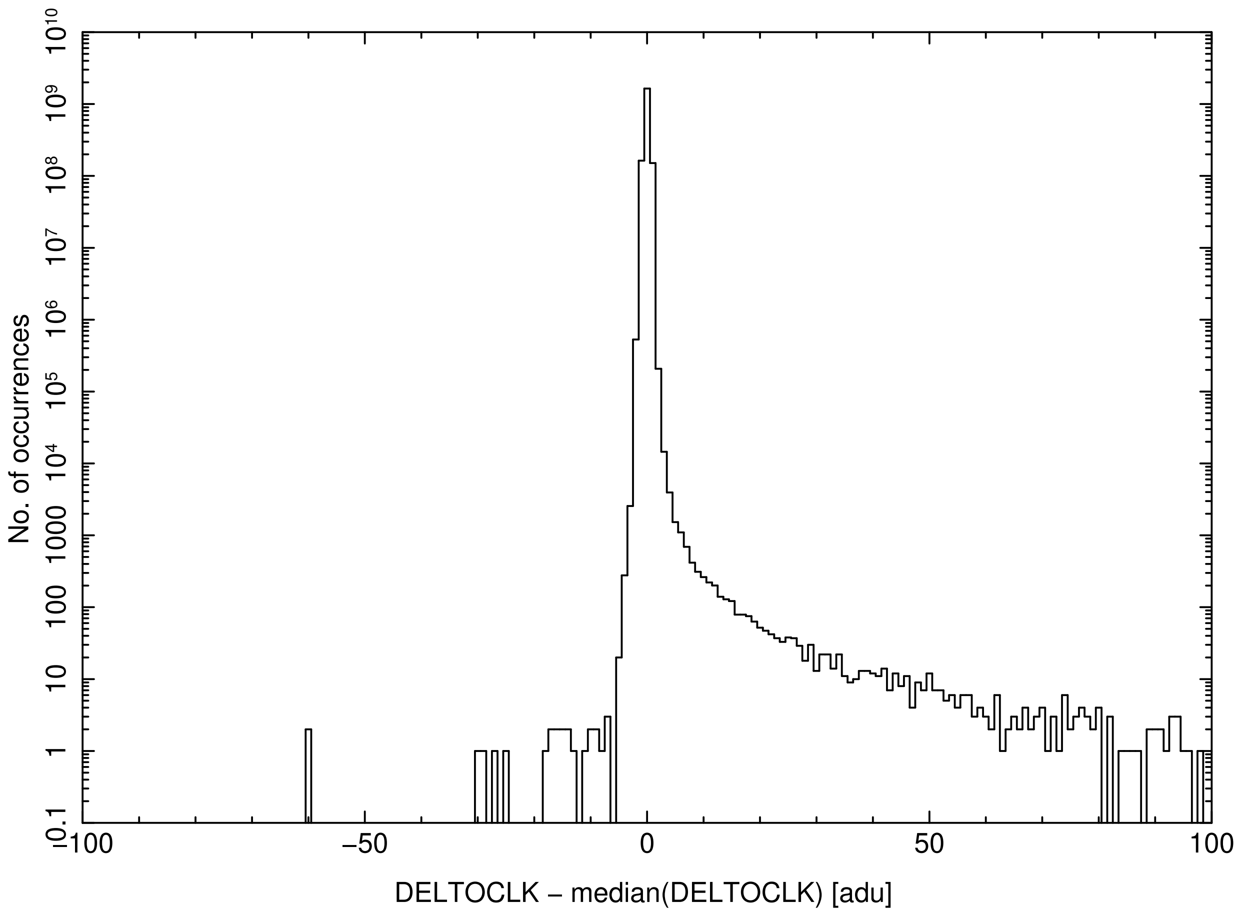

Figure 4: This figure is identical to Figure 3, except that it includes only the data in the range from -100 to 100 adu.

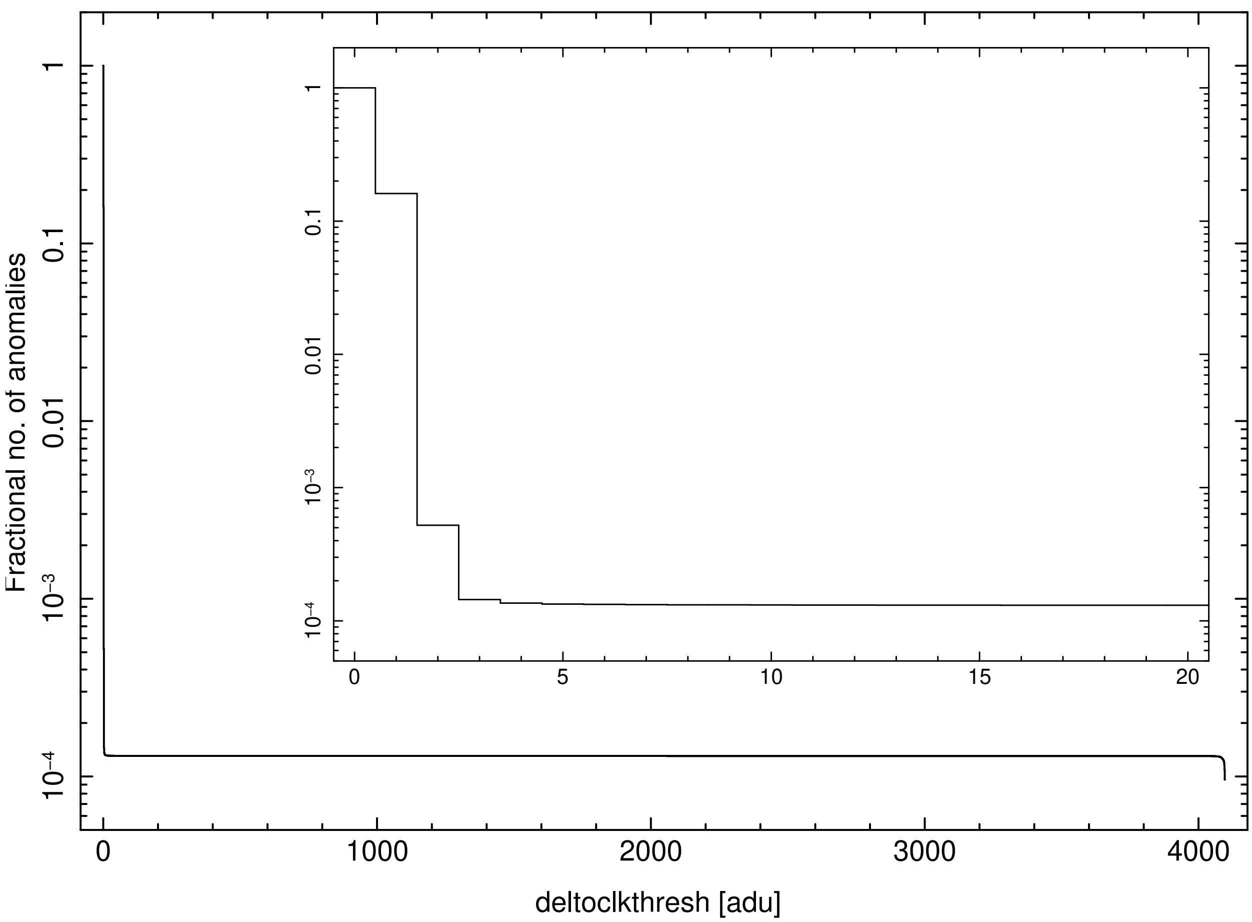

Figure 5: A plot of the fraction of the values of `DELTOCLK` that are identified as anomalous as a function of the parameter `deltoclkthresh`. This plot was produced using the data that is shown in Figure 2. The inset histogram is a plot of the same data, except that it includes only the data in the range from 0 to 20.