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

MEMORANDUM
February 19, 2013

To:	Jonathan McDowell, SDS Group Leader
From:	Glenn E. Allen, SDS
Subject:	`DELTOCLK spec`
Revision:	3.1
URL:	http://space.mit.edu/CXC/docs/docs.html#deltoclk
File:	/nfs/inconceivable/d0/SDS/SPECS/DELTOCLK/deltoclk_spec_3.1.tex

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 of a particular CCD (e.g. Fig. 1). These anomalies are primarily associated with the front-illuminated CCDs (Fig. 2). The pulse-height data associated with such anomalies are inaccurate because

$$\texttt{PHAS}[j,i,k] = \texttt{RAW\_PHAS}[j,i,k] - \texttt{BIAS}[j,i] - \texttt{DELTOCLK}[k]. \quad (1)$$

In equation 1, RAW_PHAS$[j,i,k]$ is the unadjusted pulse height read out by an ACIS detector for pixel $[j,i]$[*] of an event island in the frame where EXPNO $= k$, BIAS$[j,i]$ is the bias value associated with the pixel $[j,i]$, DELTOCLK$[k]$ is the value of DELTOCLK for frame k of the node that includes pixel $[j,i]$,[†] and PHAS$[j,i,k]$ is the bias- and DELTOCLK-adjusted pulse height.

This memo describes how to identify and handle anomalous values of DELTOCLK. The identification algorithm accommodates the following features: (1) The initial values of DELTOCLK are typically invalid (e.g. 4095, Fig. 3 and Table 1). (2) There can be a large positive or negative gradient in the values of DELTOCLK at the beginning of an observation until the temperature of a detector stabilizes (Fig. 4). (3) Periodic variations with amplitudes of a few adu occur in at least some of the data (Fig. 5).

The identification algorithm does not handle the rare case where an input `exrfile` has one or more "frame gaps" (Fig. 6). Here, a frame gap is defined as a case where there is no DELTOCLK data for several consecutive frames (i.e. EXPNOs). An examination of the data suggests that the sparsely sampled values of DELTOCLK associated with such gaps are not anomalous and that the well-sampled data immediately preceding and immediately following a gap are handled well. Therefore, frame gaps do not seem to be a significant problem.

Once anomalies are identified, Level 1 event-data files are modified to set a STATUS bit to one for events that are adversely affected by anomalies and, if possible, to adjust the pulse-height data to compensate for

[*] For a FAINT-mode event that occurs on the pixel (CHIPX, CHIPY) $= (x, y)$, the indices $i = 0$, 1, and 2 correspond to CHIPX $= x-1$, x, and $x+1$, respectively, and the indices $j = 0$, 1, and 2 correspond to CHIPY $= y-1$, y, and $y+1$. Similarly, for a VFAINT-mode event, the indices $i = 0$, 1, 2, 3, and 4 correspond to CHIPX $= x-2$, $x-1$, x, $x+1$, and $x+2$ and the indices $j = 0$, 1, 2, 3, and 4 correspond to CHIPY $= y-2$, $y-1$, y, $y+1$, and $y+2$.

[†] The NODE_ID = 0, 1, 2, or 3 for CHIPX = 1–256, 257–512, 513–768, or 769-1024, respectively.

the anomalies. Events that have the STATUS bit set are excluded from Level 2 event files. Level 1 exposure-statistics files are modified to adjust the overclock data to compensate for anomalies and to include a list of the anomalies. The data in such lists could be used to compensate for any changes to ARFs and instrument maps (see sec. 6). Level 0 data files remain unchanged (i.e. include the anomalies) so that it is possible to reprocess the data with a different algorithm.

2 Input

1. The EXPNO and DELTOCLK data in the EXR HDU of one or more Level 0 exposure-records file(s) [exrfile(s)]

2. The EXPNO and RAW_PHAS data in the EVENTS HDU of one or more Level 0 event-data file(s) [infile(s)]

3 Output

1. For CC33_FAINT, FAINT, FAINT_BIAS, or VFAINT mode observations, but not CC33_GRADED or GRADED mode observations, the values of PHAS (and, subsequently, ENERGY, FLTGRADE, GRADE, PHA, PHA_RO, and PI) in the EVENTS HDU of the Level 1 event-data file (outfile) are updated to compensate for DELTOCLK anomalies.

2. The STATUS bit information in the EVENTS HDU of the Level 1 event-data file (outfile) is updated to indicate which events are affected by anomalies.

3. The values of OVRCLOCK in the EXPSTATS HDU of the Level 1 exposure-statistics file (expstatfile) are adjusted to compensate for DELTOCLK anomalies.

4. A BADOCLK HDU is created in the Level 1 exposure-statics file (expstatfile).

4 Parameters

1. `infile,f,a,"",,,"Name(s) of input Level 0 event-data file(s)"`

2. `outfile,f,a"",,,"Name of output Level 1 event-data file"`

3. `exrfile,f,a,"",,,"Name(s) of input Level 0 exposure-records file(s)"`

4. `expstatfile,f,a,"",,,"Name of output Level 1 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"`

   The default value of maxframegap is 11 because the value of DTYCYCLE has been as large as 10 and because consecutive frame numbers typically differ by eleven or less (Fig. 7).

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"`

   The default value of deltoclkthresh is 3 because a value of three seems to represent a good balance between maximizing the number of anomalies identified and minimizing the rate of false positives (Fig. 8).

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`(s) 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`.

2. To determine if one or more values of `DELTOCLK` is anomalous for a node, the data in an `exrfile` are processed to compute the values M for the node. As described below, the values M are estimates of what the values of `DELTOCLK` would be in the absence of anomalies. Except in special circumstances (see below), the estimates M are obtained by smoothing the data with a sliding median filter.

   (a) For the first row of data, the value of M_1 is determined as follows. Note that the subscript denotes the row number in the `exrfile`, not the frame number.

      i. If there are at least two rows of data in the `exrfile` and if $\texttt{EXPNO}_2 = \texttt{EXPNO}_1 + 1$, then $M_1 = \texttt{DELTOCLK}_2$.

      ii. If there is only one row of data in the `exrfile` or if $\texttt{EXPNO}_2 \neq \texttt{EXPNO}_1 + 1$, then

$$M_1 = \texttt{DELTOCLK}_1 \quad \text{if } \texttt{DELTOCLK}_1 \neq 4095 \tag{2}$$

$$M_1 = 0 \quad \text{if } \texttt{DELTOCLK}_1 = 4095 \tag{3}$$

   (b) If there are at least two rows of data in the `exrfile`, then the value of $M_2 = \texttt{DELTOCLK}_2$.

   (c) If there are fewer than three rows of data in the `exrfile` or if the value of $n = 1$ (the smallest possible value for n), then skip step 2c. Here,

$$n = \frac{\texttt{numframes} - 1}{2}. \tag{4}$$

   If $n \geq 2$ and there are at least three rows of data in the `exrfile`, then for rows r from 3 to $n+1$, the value of M_r is the median[‡] of the set [$\texttt{DELTOCLK}_{r-1}$, $\texttt{DELTOCLK}_r$, $\texttt{DELTOCLK}_{r+1}$], provided

$$\texttt{EXPNO}_r - \texttt{EXPNO}_{r-1} \leq \texttt{maxframegap} \text{ and} \tag{5}$$

$$\texttt{EXPNO}_{r+1} - \texttt{EXPNO}_r \leq \texttt{maxframegap}. \tag{6}$$

   If either equation 5 or 6 is not satisfied, then $M_r = \texttt{DELTOCLK}_r$ because there is too little data to accurately estimate M_r (i.e. to determine if the value of $\texttt{DELTOCLK}_r$ is anomalous). Note that the data in the rows from 3 to $n+1$ are handled using an effective value of `numframes` of 3 since there can be a relatively large gradient in the values of `DELTOCLK` in the first few rows of data.

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

$$\texttt{EXPNO}_{r-1} - \texttt{EXPNO}_{r-2} \leq \texttt{maxframegap}, \tag{7}$$

$$\texttt{EXPNO}_r - \texttt{EXPNO}_{r-1} \leq \texttt{maxframegap}, \tag{8}$$

$$\texttt{EXPNO}_{r+1} - \texttt{EXPNO}_r \leq \texttt{maxframegap}, \text{ and} \tag{9}$$

$$\texttt{EXPNO}_{r+2} - \texttt{EXPNO}_{r+1} \leq \texttt{maxframegap}. \tag{10}$$

[‡] 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 medians of the sorted sets [1,1,4] and [1,1,4,5,6] are 1 and 4, respectively. 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.

If equation 7 is not satisfied, then $\texttt{DELTOCLK}_{r-2}$ is excluded from the computation of the median because the frame gap is too large. If equation 8 is not satisfied, then both $\texttt{DELTOCLK}_{r-2}$ and $\texttt{DELTOCLK}_{r-1}$ are excluded from the computation of the median. If equation 9 is not satisfied or if row r is the last row, then both $\texttt{DELTOCLK}_{r+1}$ and $\texttt{DELTOCLK}_{r+2}$ are excluded from the computation of the median. If equation 10 is not satisfied or if row $r+1$ is the last row, then $\texttt{DELTOCLK}_{r+2}$ is excluded from the computation of the median. If there are not at least `minnumframes` values included in the computation of the median, then $M_r = \texttt{DELTOCLK}_r$ because there is too little data to accurately estimate M_r.

(e) Steps 2c and 2d, but not steps 2a and 2b, are repeated an additional `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_r is the median of the previous set of values $[M_{r-n}, \ldots, M_{r+n}]$ instead of the median of $[\texttt{DELTOCLK}_{r-n}, \ldots, \texttt{DELTOCLK}_{r+n}]$. 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_r 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_r where there are more data points in the computation on one side of frame r 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).

3. Once the estimates M are computed for a node, the values of DELTOCLK for the node are compared to the values of M to determine if one or more value of DELTOCLK is anomalous. The value of DELTOCLK for frame k is identified as anomalous if

$$\texttt{DELTOCLK}[k] \geq M[k] + \texttt{deltoclkthresh} \text{ or} \quad (11)$$

$$\texttt{DELTOCLK}[k] \leq M[k] - \texttt{deltoclkthresh}. \quad (12)$$

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

$$\texttt{PHAS}[j,i,k] = \texttt{RAW\_PHAS}[j,i,k] - \texttt{BIAS}[j,i] - M[k] \quad (13)$$

instead of

$$\texttt{PHAS}[j,i,k] = \texttt{RAW\_PHAS}[j,i,k] - \texttt{BIAS}[j,i] - \texttt{DELTOCLK}[k] \quad (14)$$

for all of the events that occur during frame k on the particular CCD and node with an anomaly.[§] If an event occurs along the edge of a node that has an anomalous value of DELTOCLK, then only the pulse-height values $[j,i]$ 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 occurs on a good node that is adjacent to a node with an anomalous value of DELTOCLK, then only the pulse heights of pixels $[j,i]$ that lie on the adjacent, bad node are adjusted. The pulse-height computation described by equation 13 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). This action is performed for all DATAMODEs.

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

[§] Since it is possible for a 5 pixel × 5 pixel VFAINT-mode event island to extend beyond the edges of a CCD in the CHIPX direction (but not the CHIPY direction), some care should be used when computing the values of PHAS for VFAINT-mode event data. Specifically, if DATAMODE = VFAINT and if $i = 0$ corresponds to CHIPX = 0, then $\texttt{PHAS}[j,0,k] = 0$. Likewise, if DATAMODE = VFAINT and if $i = 4$ corresponds to CHIPX = 1025, then $\texttt{PHAS}[j,4,k] = 0$.

i. The OVRCLOCKs are computed using

$$\texttt{OVRCLOCK}[k] = M[k] + \texttt{INITOCL}x \tag{15}$$

instead of

$$\texttt{OVRCLOCK}[k] = \texttt{DELTOCLK}[k] + \texttt{INITOCL}x, \tag{16}$$

where $x =$ A, B, C, or D for NODE_ID = 0, 1, 2, or 3, respectively. This action is performed for all DATAMODEs.

ii. A "BADOCLK" HDU is included, which contains a list of the EXPNO, CCD_ID, and NODE_ID information for every anomalous value of DELTOCLK. For example, if NODE_ID = 0 of CCD_ID = 7 has an anomaly for frame EXPNO = 47, then the BADOCLK HDU includes an entry for EXPNO = 47, CCD_ID = 7, and NODE_ID = 0. The other nodes of this CCD and the other CCDs for this frame are not included in the BADOCLK HDU unless they also have anomalous values of DELTOCLK. This action is performed for all DATAMODEs.

5. Steps 2–4 are performed separately for each node of each CCD.

6 Caveats

1. As of October 14, 2012, the tools **mkarf**, **mkgarf**, **mkinstmap**, and **mkwarf** have not been modified to include the time dependence introduced by specific nodes of specific CCDs 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).

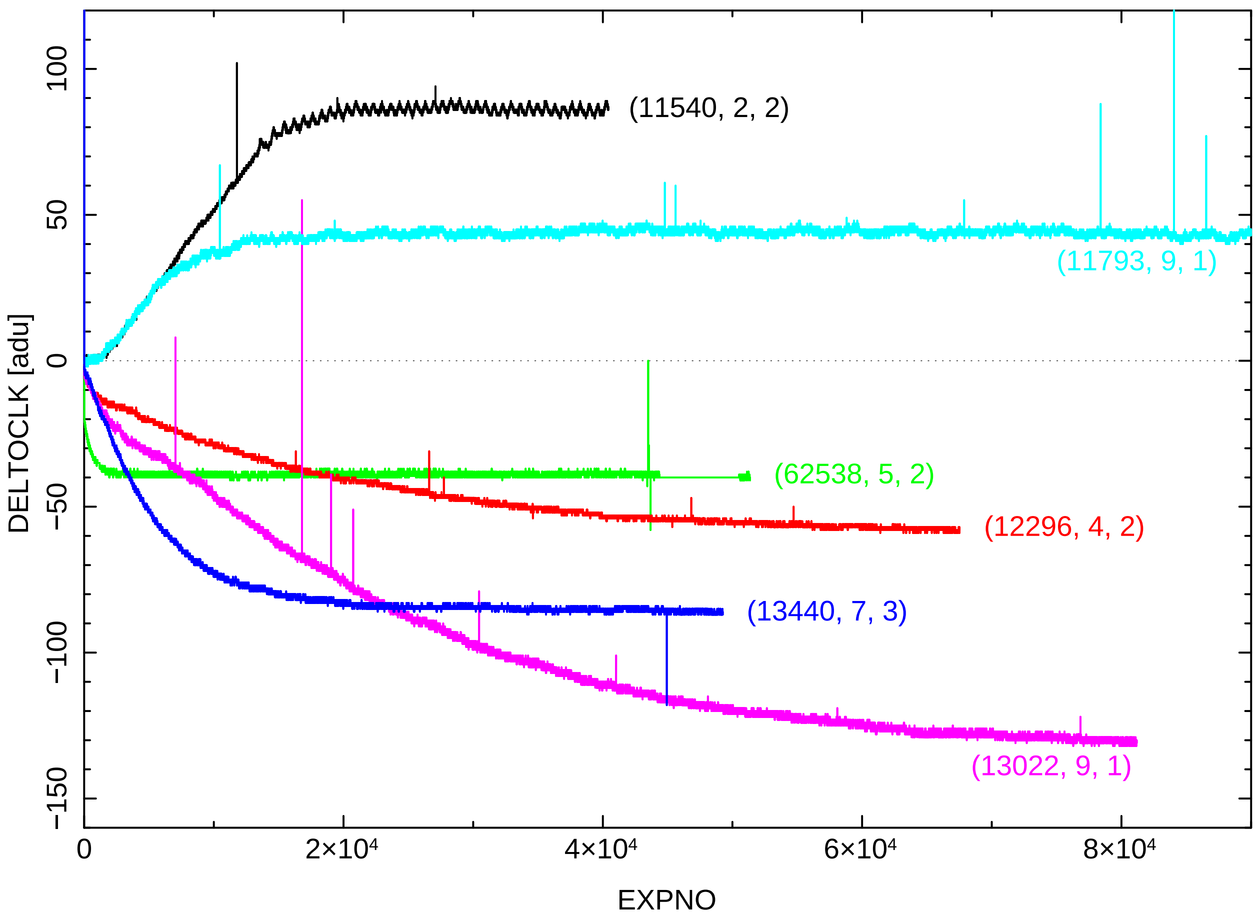

Figure 1: Examples of observations that have anomalously large and/or anomalously small values of `DELTOCLK` (i.e. the positive and negative "spikes," respectively). The sets of numbers in parentheses are the (`OBS_ID`, `CCD_ID`, `NODE_ID`). Note that the data for (11793, 9, 1) has been truncated at `EXPNO` = 90000 and at `DELTOCLK` = 120 adu.

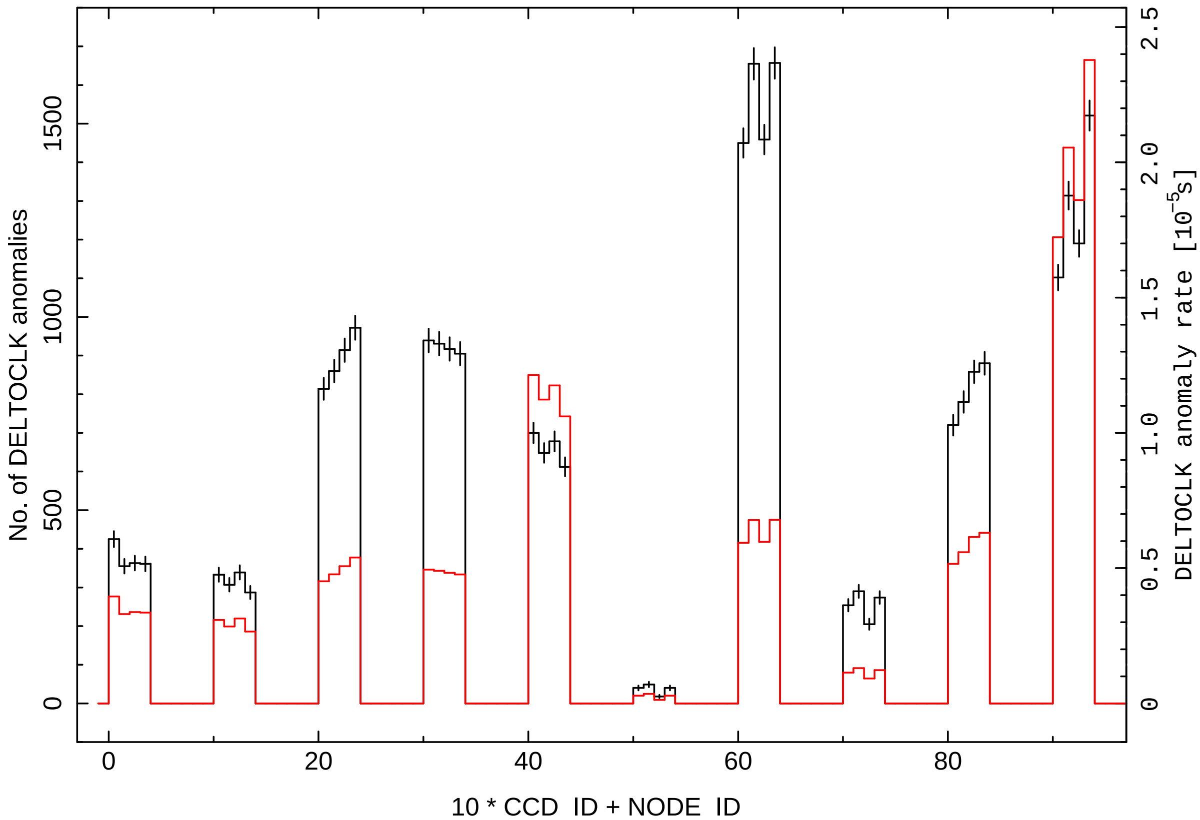

Figure 2: The black histogram (and left-hand vertical axis) is the total number of `DELTOCLK` anomalies per node for 69,899 exposure-records files obtained between July 27, 1999 and September 3, 2012. These numbers were computed using $\texttt{deltoclkthresh} = 3$. The red histogram (and right-hand vertical axis) is the corresponding anomaly rate. The time intervals for the rates were computed using the values of the keywords `TSTART` and `TSTOP` in the exposure-records files.

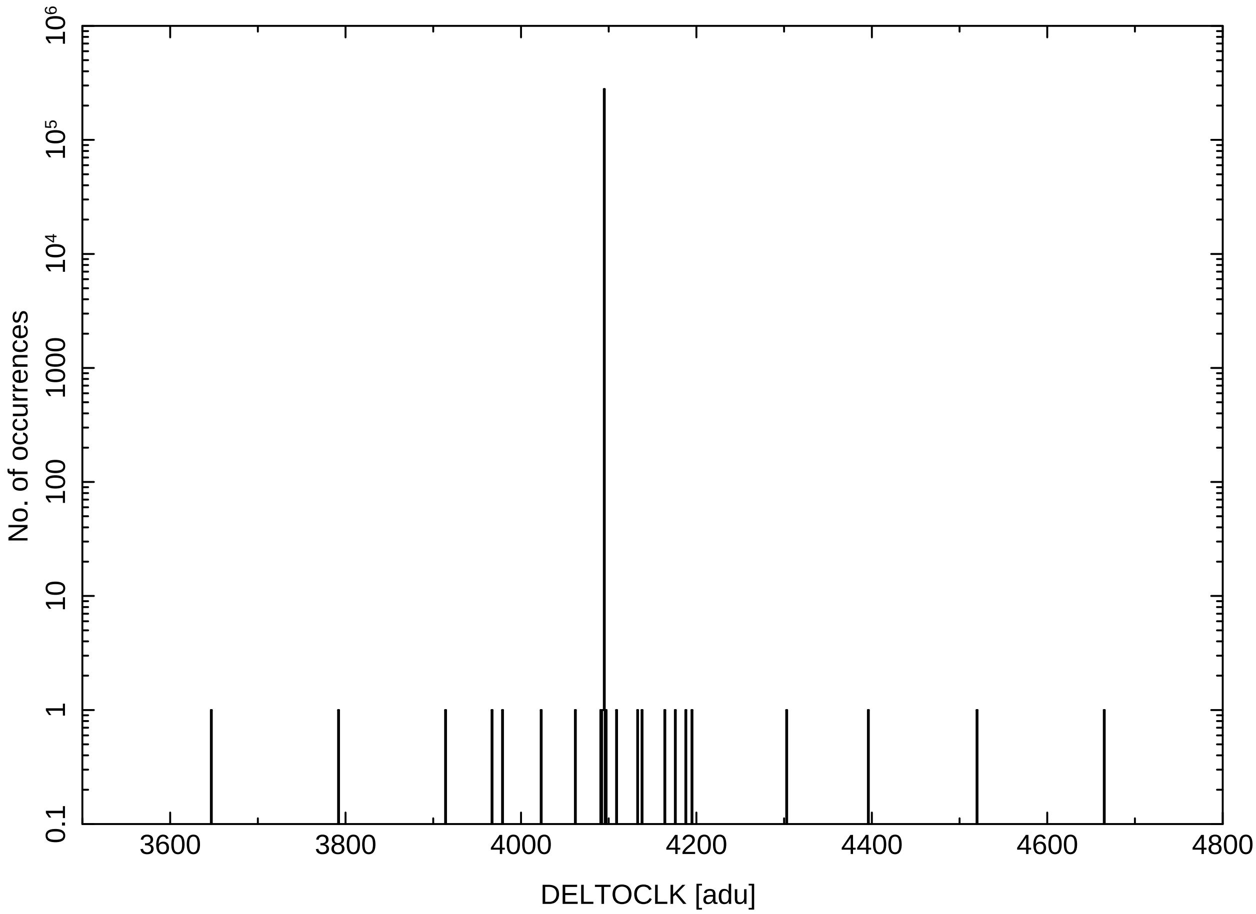

Figure 3: The distribution of the values of `DELTOCLK` in the first row of data in each one of 69,899 exposure-records files obtained between July 27, 1999 and September 3, 2012. The vast majority of these values are 4095 adu. Note that the 356 values at 0 adu (see Table 1) are not shown.

Table 1. Initial values of DELTOCLK

`DELTOCLK`	No. of occurrences
0	356
3647	1
3792	1
3914	1
3967	1
3979	1
4023	1
4062	1
4091	1
4093	1
4094	1
4095	279218
4097	1
4109	1
4133	1
4138	1
4164	1
4176	1
4188	1
4195	1
4303	1
4396	1
4520	1
4665	1
Total	279596

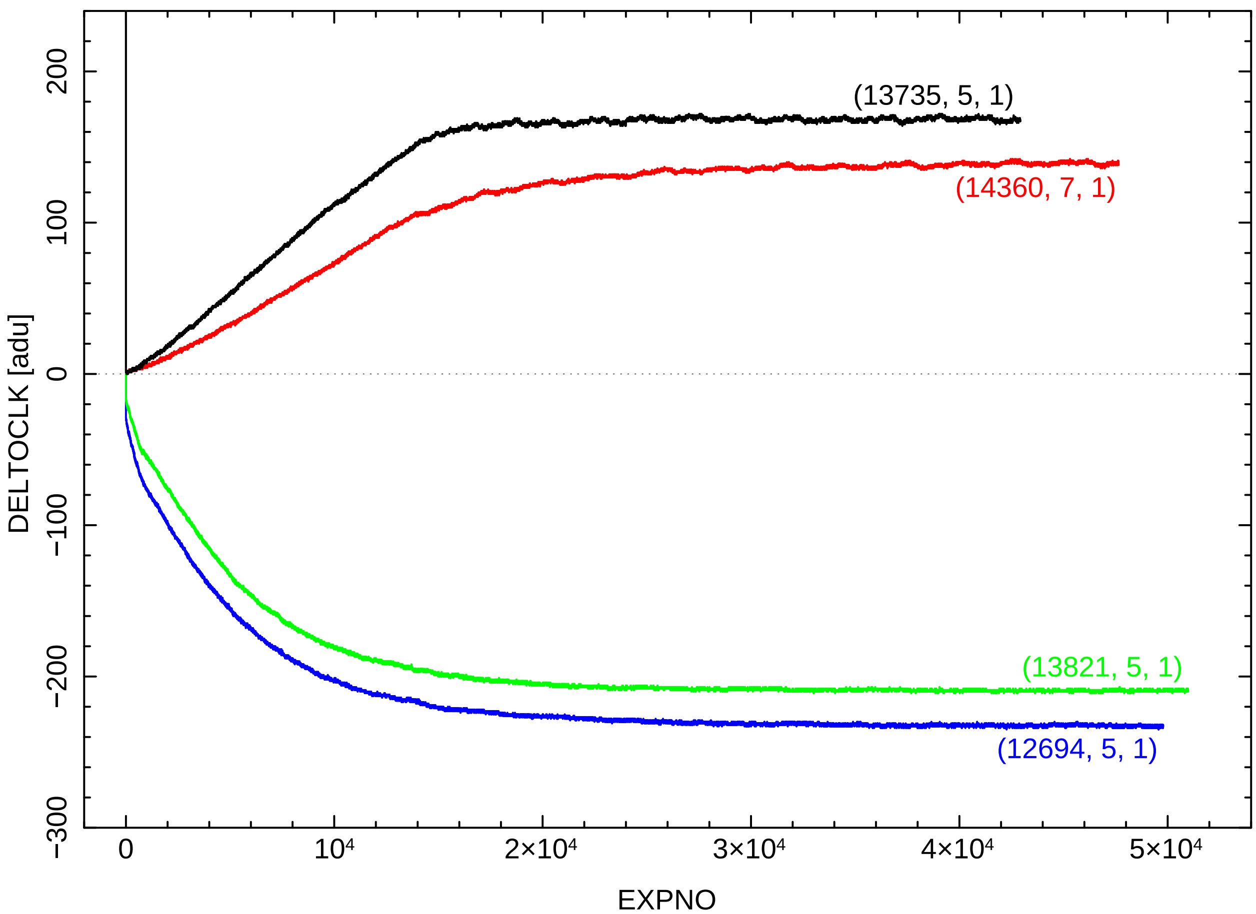

Figure 4: Examples of observations that have a large positive or negative gradient in the values of DELTOCLK at the beginning of the observation. The sets of numbers in parentheses are the (OBS_ID, CCD_ID, NODE_ID).

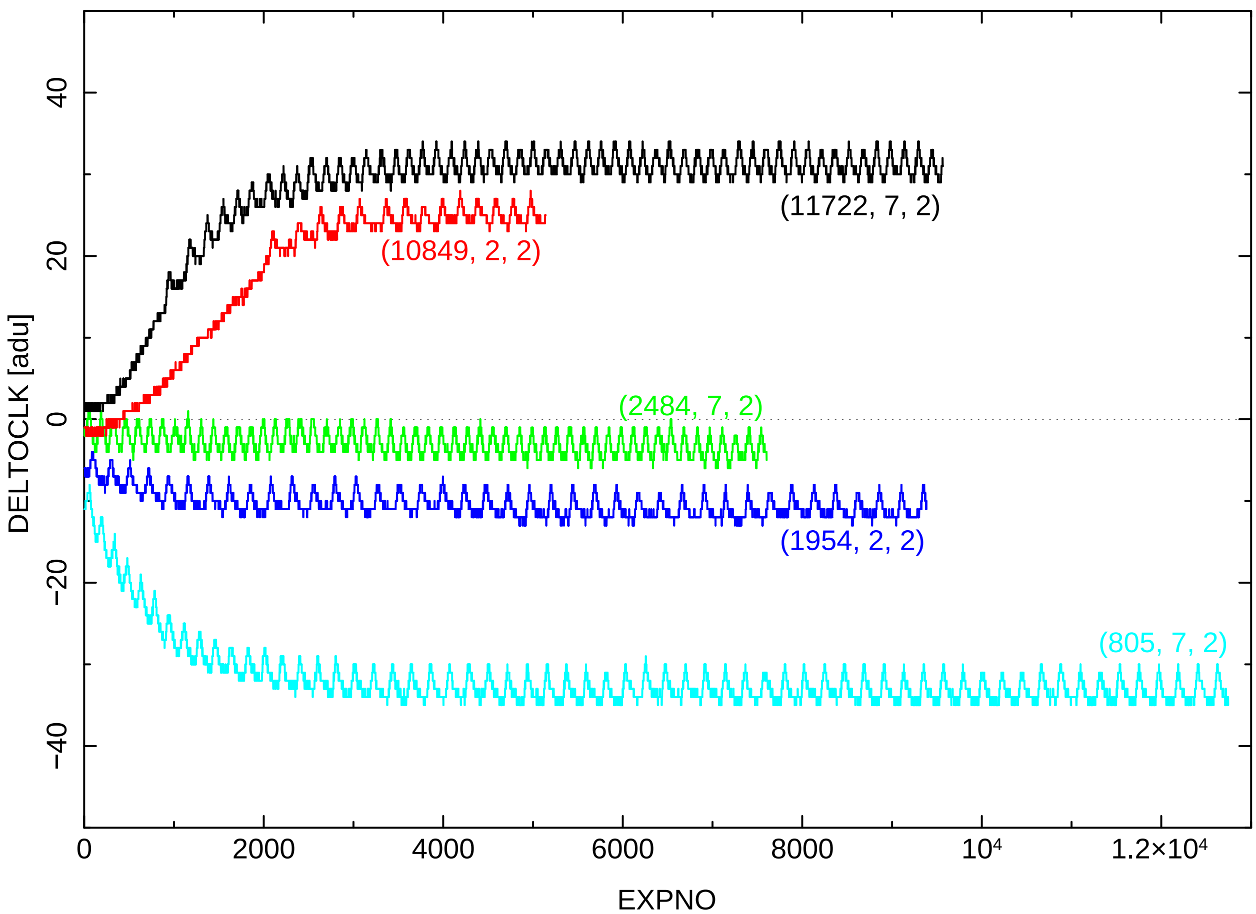

Figure 5: Examples of observations that have periodic variations in the values of DELTOCLK. The sets of numbers in parentheses are the (OBS_ID, CCD_ID, NODE_ID). An offset of +5 adu was added to the values of DELTOCLK for (11722, 7, 2) to improve visual clarity. For (10849, 2, 2) and (11722, 7, 2), the variations appear to begin after about 2000 and 900 frames, respectively. Similarly, variations appear to begin after about 13000 frames of data for (11540, 2, 2) in Figure 1.

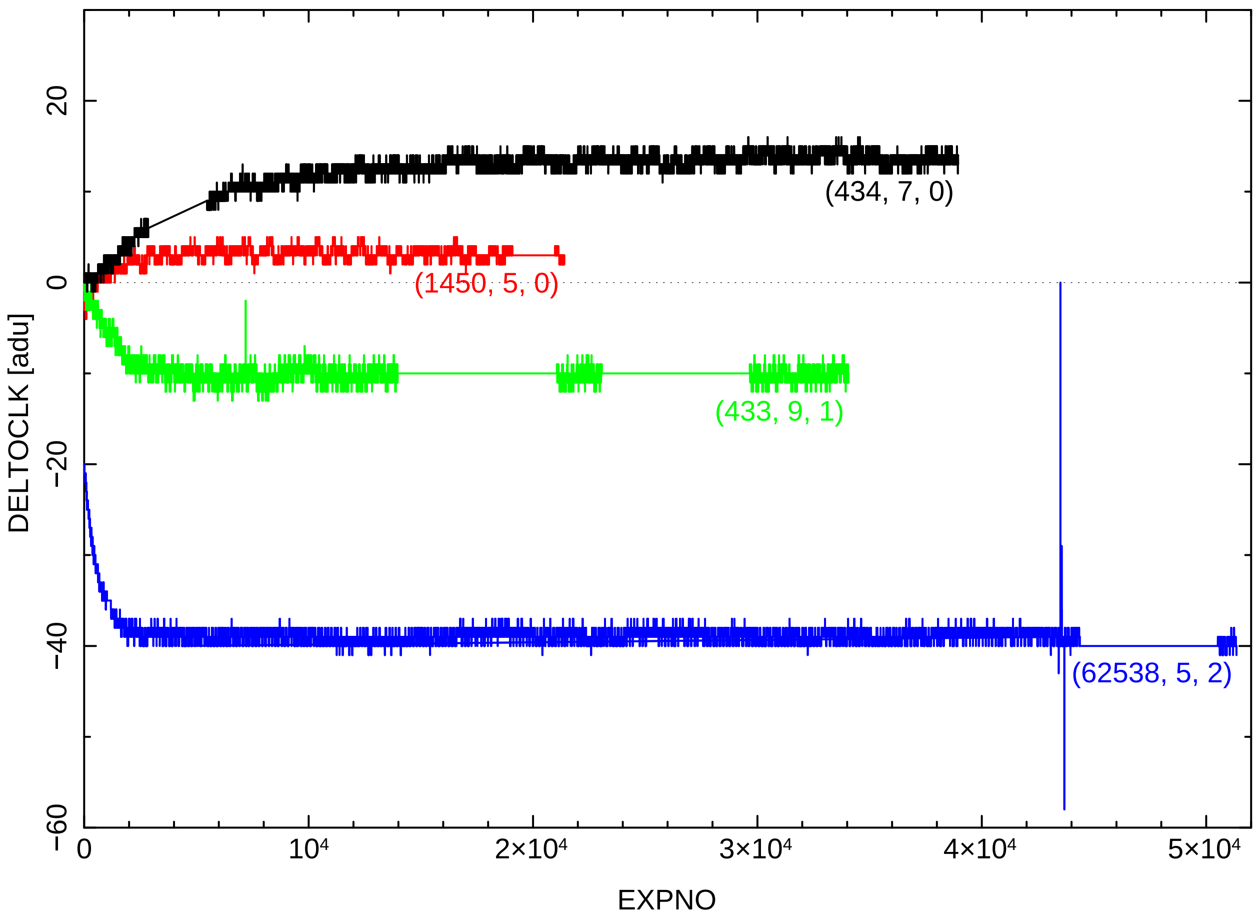

Figure 6: Examples of observations that have "frame gaps." Here, a frame gap is defined as a case where there is no `DELTOCLK` data for several consecutive frames (i.e. `EXPNO`s). The sets of numbers in parentheses are the (`OBS_ID`, `CCD_ID`, `NODE_ID`). For (433, 9, 1), there are two obvious frame gaps. For the other three cases shown, there is one obvious frame gap. Straight lines connect the data on one side of a frame gap with the data on the other side.

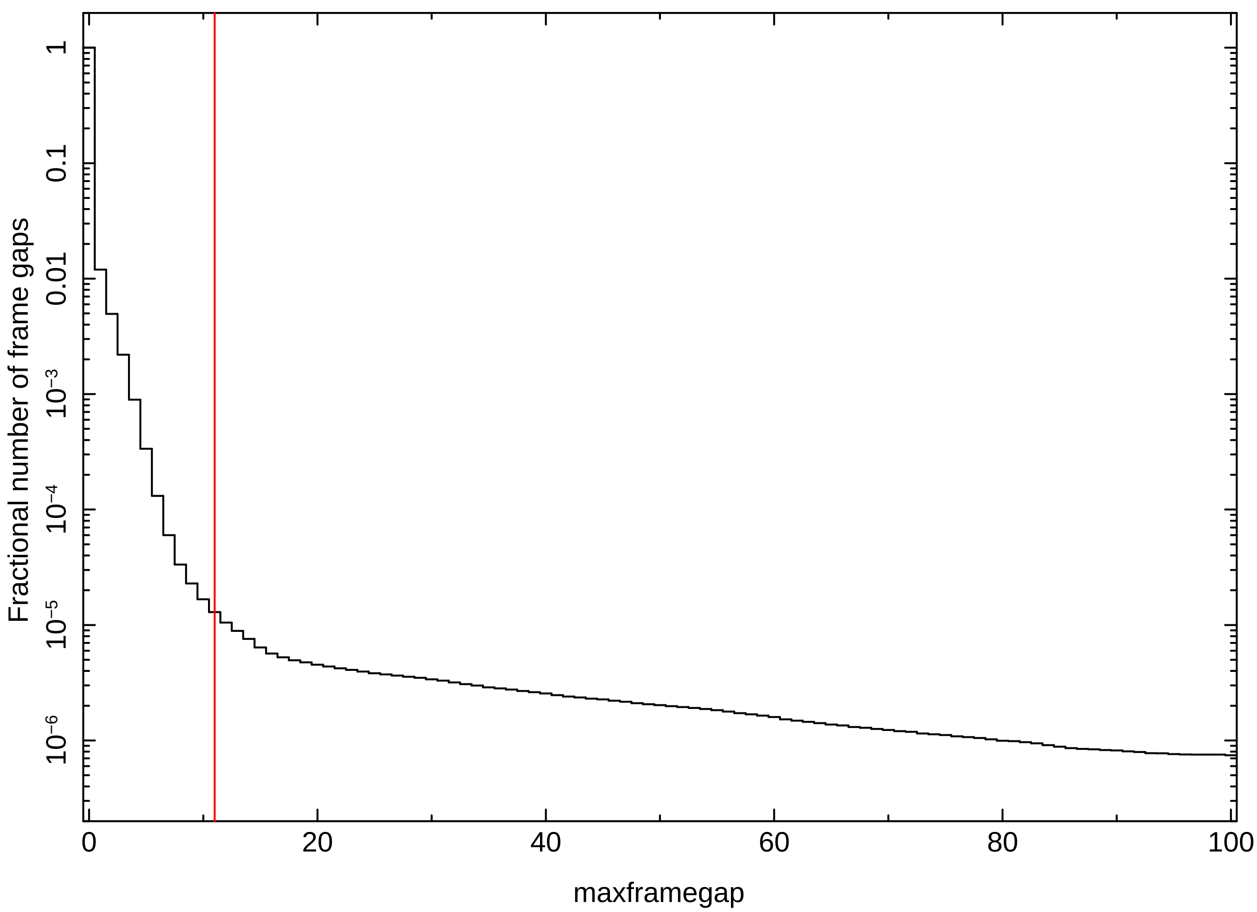

Figure 7: The fraction of the data that is identified as having a "frame gap" as a function of the parameter `maxframegap`. The vertical red line indicates that the default value of `maxframegap` is 11. For this value, about 1 of every 10^5 sets of consecutive pairs of frame numbers (i.e. `EXPNO`s) differs by more than `maxframegap` and, hence, is considered a frame gap. This plot was created using more than 5.43×10^8 pairs of consecutive `EXPNO`s from 69,899 exposure-records files obtained between July 27, 1999 and September 3, 2012.

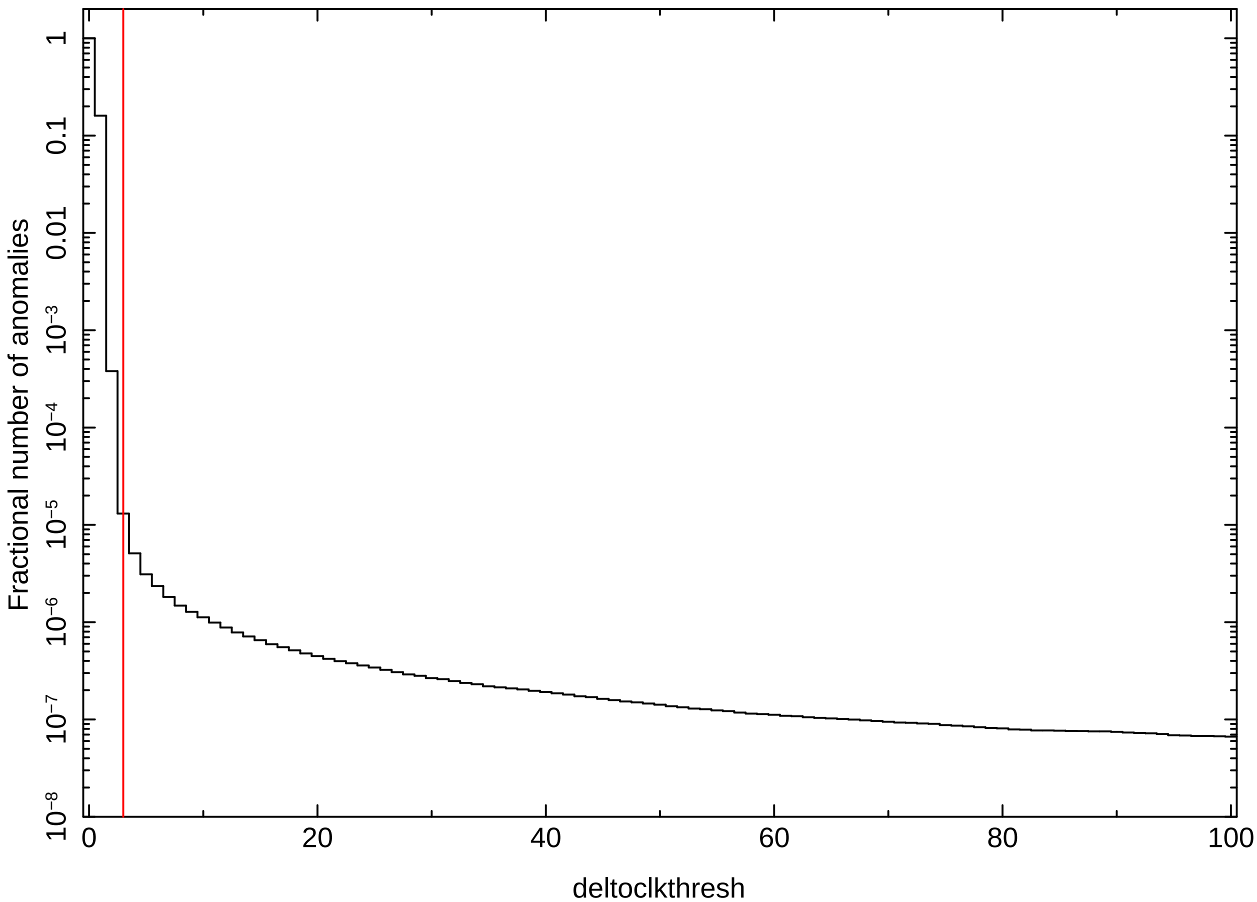

Figure 8: The fraction of the values of `DELTOCLK` that are identified as anomalous as a function of the parameter `deltoclkthresh`. The vertical red line indicates that the default value of `deltoclkthresh` is 3. For this value, about 1 in every 10^5 values of `DELTOCLK` is identified as anomalous, which represents an upper limit on the false positive rate. This plot was created using more than 2.17×10^9 values of `DELTOCLK` from 69,899 exposure-records files obtained between July 27, 1999 and September 3, 2012. Note that if the initial value of `DELTOCLK` is 4095, then this invalid value was excluded from data used to create this plot.