Pixel-Level Calibration in the *Kepler* Science Operations Center Pipeline

Elisa V. Quintana[*a], Jon M. Jenkins[a], Bruce D. Clarke[a], Hema Chandrasekaran[b], Joseph D. Twicken[a], Sean D. McCauliff[c], Miles T Cote[d], Todd C. Klaus[c], Christopher Allen[c], Douglas A. Caldwell[a], Stephen T. Bryson[d]

[a]SETI Institute, 515 N. Whisman Road, Mountain View, CA USA 94043;
[b]Lawrence Livermore National Laboratory, P.O. Box 808, L-478, Livermore, CA, USA 94551;
[c]Orbital Sciences Corporation, MS 244-30, Moffett Field, CA, USA 94035;
[d]NASA Ames Research Center, MS 244-30, Moffett Field, CA, USA 94035

ABSTRACT

We present an overview of the pixel-level calibration of flight data from the *Kepler Mission* performed within the *Kepler* Science Operations Center Science Processing Pipeline. This article describes the calibration (CAL) module, which operates on original spacecraft data to remove instrument effects and other artifacts that pollute the data. Traditional CCD data reduction is performed (removal of instrument/detector effects such as bias and dark current), in addition to pixel-level calibration (correcting for cosmic rays and variations in pixel sensitivity), *Kepler*-specific corrections (removing smear signals which result from the lack of a shutter on the photometer and correcting for distortions induced by the readout electronics), and additional operations that are needed due to the complexity and large volume of flight data. CAL operates on long (~30 min) and short (~1 min) sampled data, as well as full-frame images, and produces calibrated pixel flux time series, uncertainties, and other metrics that are used in subsequent Pipeline modules. The raw and calibrated data are also archived in the Multi-mission Archive at Space Telescope at the Space Telescope Science Institute for use by the astronomical community.

Keywords: *Kepler*, Space Telescope, Transit Photometry, Calibration

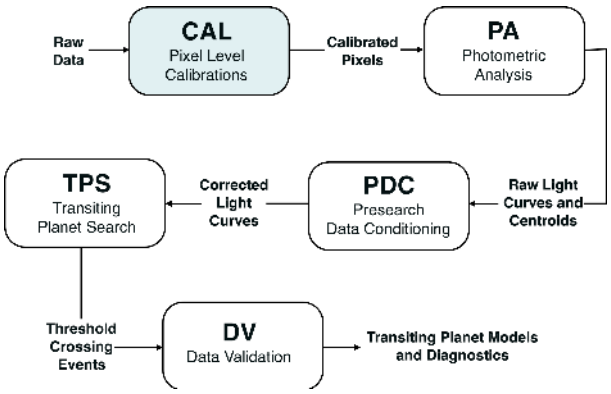
1. INTRODUCTION

The primary goal of the *Kepler Mission* is to detect Earth-size planets transiting Sun-like stars outside of the Solar System. In order to obtain the precision needed to detect such small photometric signals, the flight data is subject to numerous data processing steps that are performed in a complex automated pipeline which was developed in the Kepler Science Operations Center (SOC). The SOC Science Processing Pipeline[1,2] (hereafter referred to as the Pipeline) is composed of a number of modules that operate sequentially (Figure 1), including calibration (CAL), photometric analysis[3] (PA), pre-search data conditioning[4] (PDC), transiting planet search[5] (TPS), and data validation[6] (DV). Additional modules operate in parallel to monitor the instrument performance[7,8] and provide target management[9]. The flight data are processed monthly and quarterly[10], and are subject to continuous reprocessing throughout the length of the mission as the Pipeline algorithms continue to improve. Once each data set is processed through the Pipeline, the SOC provides a list of planetary candidates to the Science Team for further analysis.

Herein, we describe the first component of the Pipeline, CAL, which processes original flight pixel data provided by the Data Management Center at the Space Telescope Science Institute (STScI). Various data types are collected and differ by the number of integrations that compose each sampling time (or "cadence"), and also by the number and location of pixels that are collected. The raw data include photometric (target and background) pixels, along with a subset of the CCD termed "collateral data" which includes masked and virtual (over-clocked) rows and columns that are used primarily for calibration. Three types of data sets are processed within CAL: (1) select pixels from >150,000 long

[*]elisa.quintana@nasa.gov; phone 1 650 604 2467; fax 650-604-2478

cadence (LC) targets that are collected every 29.4 minutes (with 270 exposures per cadence); (2) a smaller set of pixels from 512 short cadence (SC) targets that are sampled more frequently at 0.98 minute intervals (with 9 exposures per cadence); and (3) full-frame image (FFI) data which contain all available pixels for a single long cadence. These data types are processed separately in the Pipeline, and the differences will be noted herein.



CAL	Calibrate pixels (collateral, background and target) for long cadence, short cadence, and full frame images
PA	Extract raw flux and compute photocenter (centroid) for each target and cadence from associated target pixels
PDC	Correct systematic and other errors in raw light curves, remove excess flux due to aperture crowding, and condition light curves for the transiting planet search
TPS	Perform transiting planet search and return Threshold Crossing Events (TCEs) for detections
DV	Fit transiting planet model to light curves with TCEs, search for additional transiting planets, and perform statistical tests to validate candidate planets

Figure 1. An overview of the science data flow in the SOC Pipeline is shown (left), along with a description of the primary functions of each component (right).

This article describes the data flow within CAL and the methodology and reasoning behind the individual corrections that are applied to the different pixel types. Many of the primary corrections use external models[11] of each CCD that were developed from pre-flight hardware tests and FFI data taken during commissioning[12] (prior to the dust cover ejection). We discuss how these models are applied within CAL to correct for 2D bias structure, gain and nonlinearity of the conversion from analog-to-digital units (ADU) to photoelectrons, local detector electronics effects (undershoot and overshoot), and flat field (variations in pixel sensitivity). Other signals that are corrected include excess charge from saturated stars that leak into the masked and virtual regions, cosmic ray events, dark current and smear. Note that CAL does not include any time or motion corrections or coordinate transformations. We present an overview of the focal plane CCD components and the pixels that are calibrated in the next section. Section 3 describes the individual calibration steps, presented in the order that they are performed, along with the additional functionalities of CAL. In Section 4, we present a summary of the CAL module and discuss future work that will help to improve the quality of the data.

2. KEPLER DATA FORMATS

The *Kepler* focal plane array is composed of 42 charge-coupled device (CCD) detectors (Figure 2). A CCD "module" refers to a pair of CCDs that share a field flattener and are read out simultaneously by the detector electronics. Each of the 21 modules is composed of four CCD "outputs" that are each read out by a separate analog signal chain. The CAL software component operates on a single CCD module/output, or "channel", at a time.

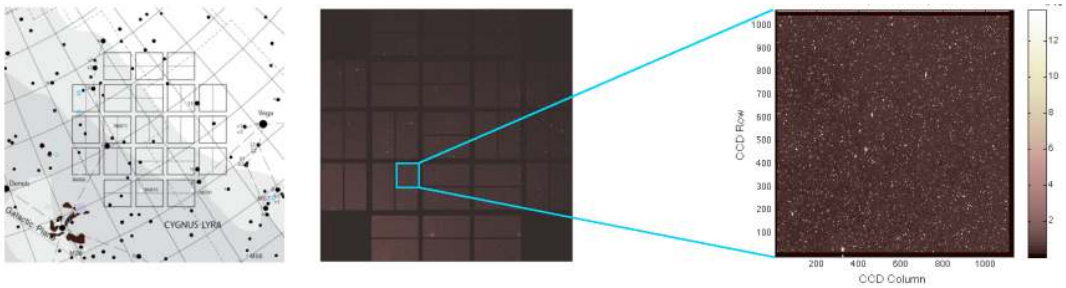


Figure 2. A celestial view of the *Kepler* focal plane (left), the first light image (middle), and a calibrated FFI from Q4 data (right, in units of 10^6 electrons/cadence) of module/output = 17/2 (channel 58) are shown.

2.1 Pixel Collection

Each CCD channel consists of an array of pixels with 1070 rows and 1132 columns, of which only a subset (1024 x 1100) is photometric (Figure 3). The full (1070 x 1132) array of pixels is downlinked for FFI data, whereas only select target and background pixels are downlinked for LC and SC data due to limitations in memory, bandwidth, and the design of the flight software[9]. For LC, an upper limit of 170,000 stellar targets and 1125 background targets are collected across the focal plane (with additional limitations on the number per channel and the total pixel count). In addition, LC collateral data (black, masked smear, and virtual smear pixels, which are described in the next section) are collected for calibration. For SC, a maximum of 512 stellar targets across the focal plane are collected, along with a subset of the collateral pixels: the black rows and smear columns that lie in the projection of the photometric pixels onto the collateral region (Figure 3), in addition to the black pixels that lie in the projection of the masked and virtual smear pixels (used to calibrate the smear).

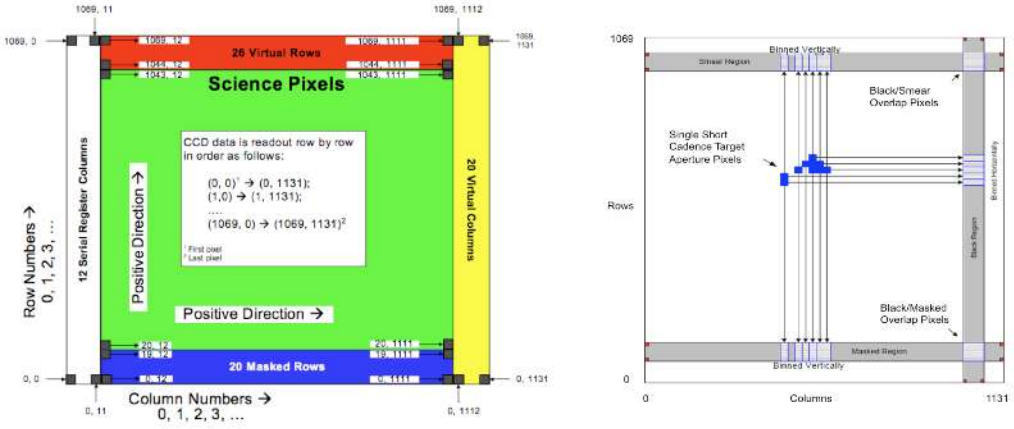


Figure 3. A schematic of the pixel regions in a single CCD channel (left) shows the location of photometric pixels, along with the collateral pixels on the perimeter of the CCD that are collected for calibration. Only a subset of collateral data (black columns and smear rows) is collected and co-added onboard the spacecraft[13]. For LC data, all black rows (and a subset of columns) and all smear columns (and a subset of rows) are collected (gray region in right panel), whereas for SC only collateral data in the projection of pixels are collected.

2.2 Photometric and Collateral Data

The photometric pixels include all available target and background pixels on a CCD channel. In PA (the Pipeline module that follows CAL), the photometric pixels are packaged into individual stellar and background targets to create target flux time series. Within CAL, however, the target and background pixels are indistinguishable and the calibration steps are processed on the individual pixel flux time series.

The collateral data includes the following: 12 "leading black" pixels that represent virtual (non-physical) pixels, which are read out before the photometric pixels in each row; 20 "trailing black" pixels, which are read out after the photometric pixels in each row; 20 "masked smear" pixels, which are physical pixels closest to the serial register that are covered with an opaque aluminum mask; and 26 "virtual smear" pixels, which are read out after all of the photometric pixels are clocked out. Only a subset of these pixels are downlinked for calibration, however, and are designated by the ground segment in a configuration map that is uplinked to the spacecraft. To estimate the black level correction, a subset of the trailing black region (columns 1119-1132) is collected, and the pixels in each row are co-added onboard the spacecraft prior to downlink. Each 'black pixel' that CAL receives is therefore the sum of 14 black pixel values per row per cadence. The leading black pixels are not used for calibration due to the presence of image artifacts in those regions. Likewise, a subset of masked smear (rows 7-18) and virtual smear (rows 1047-1058) pixels is collected, resulting in 12 co-added masked smear and 12 co-added virtual smear pixels per column per cadence. For SC data, only a small number of targets are collected from each channel, and the rows/columns of each target determine which collateral

pixels are collected. Two additional pixel types are collected for SC: “masked black” pixels, which are the sum of the pixels in the cross-sections of trailing black columns and masked smear rows, and “virtual black” pixels which are the sum of the overlapping virtual smear rows and trailing black columns. Each masked black or virtual black pixel is the sum of the number of black pixels (14) times the number of smear pixels (12), or 168 co-adds per cadence. For FFI data, all pixels are downlinked, but CAL uses the spacecraft configuration map to determine which collateral pixels should be used in calibration, and the data is processed as if it were a single long cadence.

2.3 Processing order

Data for each CCD channel are calibrated individually. Regardless of cadence type, the collateral pixels are always processed first to estimate the bias, smear, and dark levels. The photometric pixel calibration follows, using results calculated from the collateral data output. For FFIs, the collateral regions are also processed first to obtain the black, smear, and dark level estimates, and then the entire array (collateral plus photometric) is calibrated using these values.

2.4 Data Gaps

All pixel data is accompanied by logical arrays (the same size as the pixel arrays) of spatial and temporal gaps, and only the available pixels are calibrated. In some calibration steps, such as the black and dark level estimation, CAL interpolates across missing cadences. The only cases for which data may be gapped *within* CAL include (1) cadences that occur during a momentum dump, (2) cadences that occur when the spacecraft is not in fine point (when the spacecraft attitude is not precisely controlled by the fine guidance sensors[12]), and (3) masked or virtual columns that have excess flux due to saturated stars that bleeds into those columns.

3. CALIBRATION

A schematic of the data flow in the CAL module is shown in Figure 4, and the primary calibration steps are described in this section. The boxes in Figure 4 with dashed lines show the steps that can be disabled in the Pipeline if desired. All cadence types (FFI, LC, and SC) and pixel types (collateral and photometric) are processed separately, but use the same MATLAB code base (with the exceptions noted in this section). For each cadence type and channel, the first invocation processes collateral (black and smear) pixels for all cadences. The outputs to this first pass include calibrated black and smear pixels, collateral and cosmic ray metrics, and more importantly the estimates for black, smear, and dark current for the specific channel that are needed to calibrate the photometric pixels. Due to the large volume of data, the photometric pixels are subdivided by rows for LC processing, since the undershoot correction operates on pixel rows. For SC data, there are typically only 2-5 SC targets on a given channel but far more (30x) cadences than LC data, so we divide the pixels into cadence chunks.

3.1 Models

Some of the calibration steps rely on external models that characterize each CCD. These focal plane characteristics (FC) models[11] were developed during extensive ground-based tests, were updated in flight while the spacecraft dust cover was still in place, and are continuously monitored by the FC Pipeline module. These time-dependent models include (1) a read noise model, which gives the read noise per channel; (2) a 2D black model, which provides a 2D map of the black/bias structure per channel; (3) a gain model, which gives the ADU-to-photoelectrons conversion factor per channel; (4) a linearity model, which provides a set of polynomial coefficients used to correct for any nonlinearity in the gain transfer function; (5) an undershoot model, which includes filter coefficients that are used to correct for undershoot/overshoot artifacts induced by the CCD local detector electronics (LDE); and (6) a flat field model consisting of a 2D map of values that are used to correct for pixel-to-pixel sensitivity.

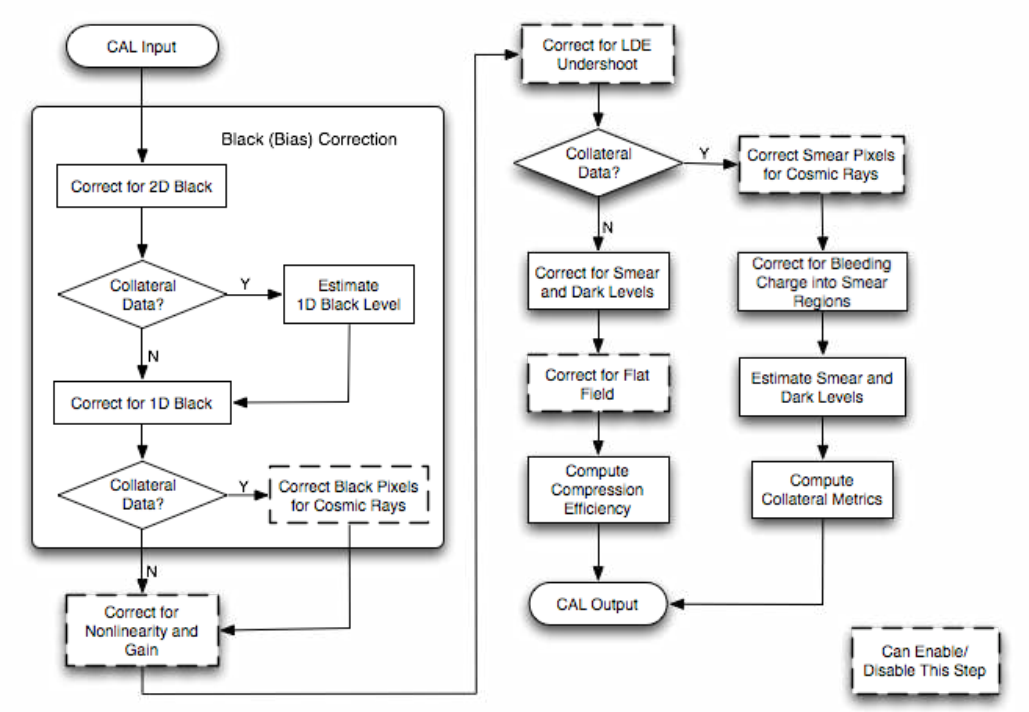


Figure 4. Data flow for calibrating collateral and photometric pixels. Dashed boxes indicate corrections that can be disabled in the Pipeline.

3.2 Fixed Offset, Mean Black, and Spatial Co-adds

Before the pixels are calibrated for black (bias) level, they are corrected for a "fixed offset" and "mean black" value. These values (which vary for SC and LC and across channels) were introduced to deal with spatial variations in bias and gain across the focal plane array, and to address issues with the pixel requantization scheme[12]. Prior to downlink, all pixels are subject to requantization in which each pixel value is mapped to a discrete value in a pre-generated table in order to control the quantization noise (the round-off error resulting from digitizing the voltage signals) to within ¼ of the intrinsic measurement uncertainty[15]. Because collateral pixels are spatially co-added and fall on a different part of the requantization table, the mean black and fixed offset work by adjusting all channels to a common zero point to ensure proper requantization. Given a pixel array ***P*** (in this case, for either collateral or photometric pixel data), the first correction performed within CAL for the available rows (*row*), columns (*col*), and cadences (*t*) is:

$$P_{all}(row, col, t) = P_{all}(row, col, t) - (fixed\ offset) + (mean\ black(t))$$

Note that MATLAB is used for all of the CAL science algorithms, so the operations are often performed on full or partial (n_{rows} x n_{cols} x $n_{cadences}$) pixel arrays rather than looping over any particular dimension. The (*row, col, t*) notation here is meant to help the reader understand which pixels are processed in each step along with the dimensions of the pixel arrays and/or corrections. CAL only operates on the available (non-gapped) rows, columns, and cadences.

The original photometric pixels are in units of ADU/cadence. The collateral pixels, however, need to be normalized by the number of spatial co-adds to convert to the same units:

$$P_{black}(row,t) = \frac{P_{black}(row,t)}{n_{black\ cols}} \quad (LC + SC\ data)$$

$$P_{masked\ smear}(col,t) = \frac{P_{masked\ smear}(col,t)}{n_{masked\ smear\ rows}} \quad (LC + SC\ data)$$

$$P_{virtual\ smear}(col,t) = \frac{P_{virtual\ smear}(col,t)}{n_{virtual\ smear\ rows}} \quad (LC + SC\ data)$$

$$P_{masked\ black}(t) = \frac{P_{masked\ black}(t)}{n_{black\ cols} \cdot n_{masked\ smear\ rows}} \quad (SC\ data\ only)$$

$$P_{virtual\ black}(t) = \frac{P_{virtual\ black}(t)}{n_{black\ cols} \cdot n_{virtual\ smear\ rows}} \quad (SC\ data\ only)$$

3.3 Black Correction

The "black level", or bias, in each CCD channel is an electronic offset that has been added to the CCD voltage to ensure that positive signals are input into the analog-to-digital converter (ADC). In addition, the black level has a 2D structure which includes various artifacts that were discovered in each channel during ground testing of the CCDs. These features are characterized in a 2D black model developed during ground testing and with reverse-clocked images (which omit starlight). Some causes of the image artifacts include heating of the readout electronics, start of line (SOL) transients, and FGS frame transfer and parallel transfer clocking crosstalk signals that are injected into the photometric region as the image is read out[13]. Figure 5 shows an example of a 2D black model (left panel) which displays SOL features near the leading black region, and a close-up view (right panel) shows frame transfer (horizontal bands) and parallel transfer crosstalk (diagonal bands) signals. A 2D black model (in ADU/exposure) is extracted within CAL for each cadence and channel, scaled by the number of exposures, and is simply subtracted off all collateral and photometric pixels:

$$P_{all}(row,col,t) = P_{all}(row,col,t) - 2Dblack(row,col,t)$$

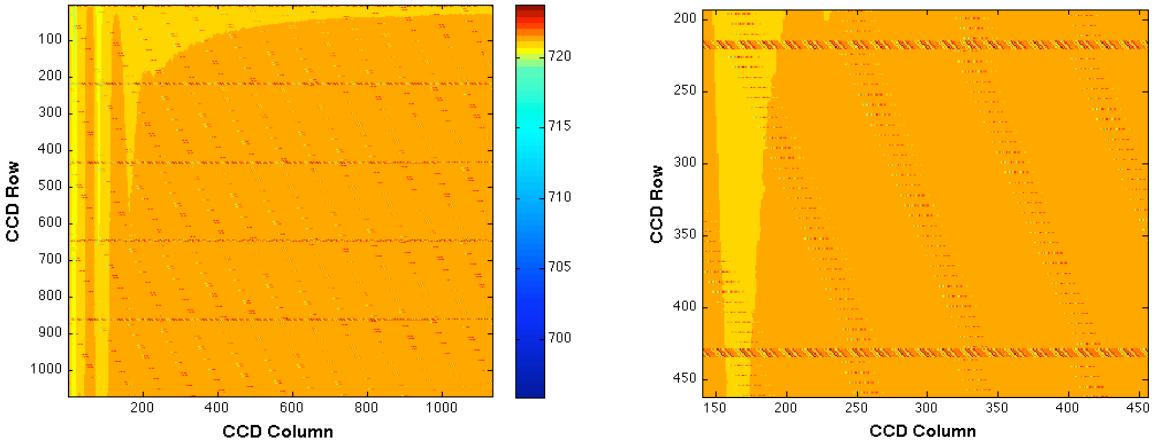


Figure 5. An example of a 2D black model (left, in units of ADU/exposure), and a close-up (right), that show the 2D bias structure that is subtracted from all pixels.

Once the 2D black level is removed, a fit to the residual bias is used to estimate a 1D black correction. The polynomial model order for the best fit is determined in an iterative fashion using the Modified Akaike Information Criterion[14]. A robust fit is first performed to protect from outliers (neglecting charge injection rows in the virtual smear region), and a least squares with known covariance method is used with the computed best polynomial order to produce the fit, or "black correction". For each cadence, the black correction in a given row is subtracted from all available pixels in that row:

$$P_{all}(row,t) \;=\; P_{all}(row,t) - 1Dblack(row,t)$$

Following the 1D black correction, the black pixels are corrected for cosmic rays and saved for output to the Multi-mission Archive at Space Telescope (MAST) at STScI, as they are no longer needed for calibration. Note that CAL only corrects LC and SC collateral pixels for cosmic rays, but uses the same methodology as the photometric pixel cosmic ray correction that is performed within PA[3].

3.4 Nonlinearity and Gain Correction

The gain and nonlinearity describe the transfer function from photoelectrons (e-) in the CCD to ADU coming out of the ADC. Gain is the average slope of the transfer function, and ranges from 94 to 120 e-/ADU across the focal plane[13,15]. Nonlinearity is a measure of the deviation from a linear transfer function at each ADU signal level. The nonlinearity model provides polynomial coefficients (Figure 6) for each exposure, and the correction can be estimated by evaluating the polynomial at the black-corrected pixel values. The range of this correction across the focal plane is within +/- 3%. The nonlinearity model is valid up until the full-well level, which is the maximum number of electrons a pixel can hold before saturation occurs (~10^6 e-). The gain model provides the gain value per channel and cadence in e-/ADU, and all pixels are simply multiplied by the gain following the nonlinearity correction. At this point, $\boldsymbol{P_{all}}$ now represents either smear arrays (in which the rows can be ignored since these are essentially just functions of columns and cadences) or photometric pixel arrays:

$$P_{all}(row,col,t) = P_{all}(row,col,t) - \; polynomial_{nonlin}(P_{all}(row,col,t))$$
$$P_{all}(row,col,t) = P_{all}(row,col,t) \cdot gain(t)$$

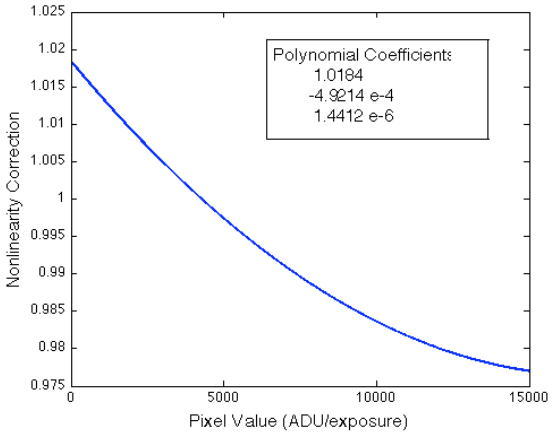


Figure 6. The nonlinearity correction, shown for a sample CCD channel, is the fractional deviation from the linear electrons-to-ADU transfer function at each pixel value.

3.5 LDE Overshoot/Undershoot Correction

Overshoot and undershoot are signal distortions that were discovered during ground testing of the CCDs, and result from operating a clamp circuit in the local detector electronics (LDE) with insufficient bandwidth[16]. The impulse response artifacts are most noticeable after light-to-dark (undershoot) and dark-to-light (overshoot) transitions, resulting in spikes in the pixel row time series of the affected targets (Figure 7). The undistorted image can be reconstructed by modeling these artifacts as a linear shift-invariant (LSI) system, which can be described by a set of difference equations that transforms an input signal ***x(n)*** into an output signal ***y(n)***:

$$\begin{aligned} a(1)\cdot y(n) \;=\; & b(1)\cdot x(n) + b(2)\cdot x(n-1) + \ldots + b(nb+1)\cdot x(n-nb) \\ & - a(2)\cdot y(n-1) - \ldots - a(na+1)\cdot y(n-na) \end{aligned}$$

Here ***n***-1 is the filter order, and ***a*** and ***b*** are the feedback and feedforward filter coefficients, respectively, that determine the z-transform system response ***H(z)***:

$$H(z) = \frac{b(1) + b(2)\cdot z^{-1} + \ldots + b(nb+1)\cdot z^{-nb}}{a(1) + a(2)\cdot z^{-1} + \ldots + a(na+1)\cdot z^{-na}}$$

The undershoot model provides a set of 20 filter coefficients for ***a***, and an inverse filter is applied (with ***b*** = 1) to each row per cadence to correct for any undershoot/overshoot:

$$P_{all}(row, t) = filter(b, a, P_{all}(row, t))$$

where *filter* is a built-in MATLAB function based on the above difference equations. Note that an extra column to the left (lower column number) of each target aperture is collected and downlinked to perform this correction.

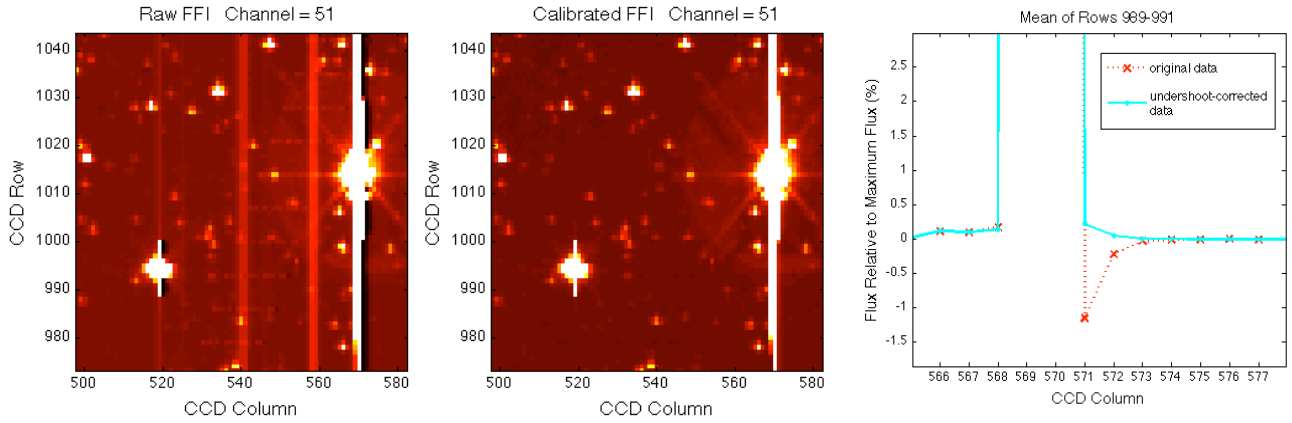


Figure 7. A close-up stretched image of two saturated target stars that show pixel undershoot signatures resulting from bright-to-dark pixel transitions in the direction of the serial readout (left panel), and the calibrated image (middle). The mean of 3 pixel rows is shown for one target (right) with the undershoot response (the negative spike) along with the corrected pixel values.

Both collateral and photometric data are corrected for undershoot/overshoot, and the median value across the focal plane array[15] is ~0.34%. In the collateral data invocation, the LC and SC masked and virtual smear pixels are next corrected

for cosmic rays and saved for output. They are still used to estimate the smear and dark current levels that are needed to correct the photometric pixels (as described in the next section), but the "calibrated smear" pixels that are output to the MAST consist of the calibrated pixels up to this point.

3.6 Smear and Dark Correction

The target and background pixels are corrected for both smear and dark current levels. The *Kepler* photometer is operated without a shutter, so stars smear along columns as the CCD is read out and are clearly visible in FFI data (Figure 8). Dark current is a thermally induced signal in each physical pixel during an integration period, which includes the exposure time ($t_{exposure}$ ~ 6.02 s) and readout time (t_{read} ~ 0.52 s). Because the focal plane is maintained at such a cold temperature (-85 degrees C), the dark current is very low with a median value of ~0.25 e-/pixel/sec across the focal plane[15]. The smear and dark corrections are grouped together here because they can be estimated from linear combinations of the virtual and masked smear pixels.

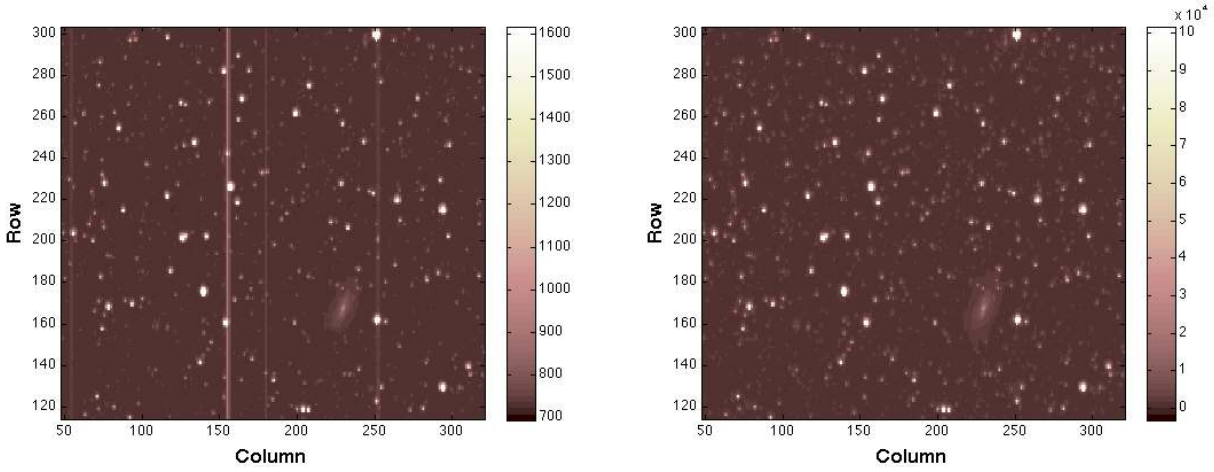


Figure 8. A portion of an uncalibrated FFI (in units of ADU/cadence, left) and the calibrated image (in photoelectrons per cadence, right) demonstrate the removal of smear from several columns.

The masked smear pixels, which are shielded from star flux, detect dark current during an integration ($t_{exposure} + t_{read}$) and collect smear signal from the photometric and over-clocked virtual pixels during readout. The virtual pixels contain dark current that is accumulated during t_{read} only, but collect smear as they are clocked through the image. The dark level per cadence is computed by taking a robust mean of the masked and virtual smear differences from the common columns:

$$dark\ level(t) = mean\left(P_{masked\ smear}(col,t) - P_{virtual\ smear}(col,t) \cdot \left(\frac{t_{\text{exposure}} + t_{\text{read}}}{t_{\text{exposure}}} \right) \right)$$

We interpolate dark level values over missing cadences to ensure that a dark level is available for all cadences. To compute the smear level, the dark level is first removed from the masked and virtual pixels. Ideally, both masked and virtual pixels are available for each column and cadence, but either may be used if only one is available. If neither is available, however, the smear correction cannot be performed for the entire column. We use the (n_{cols} x $n_{cadences}$) logical gap indicator arrays $\boldsymbol{G}$ (where gaps = true) that are provided with the smear pixel arrays to estimate the smear levels:

$$P_{masked\ smear}(col,t) = P_{masked\ smear}(col,t) - (dark\ level(t)) \cdot G_{masked\ smear}(col,t)$$

$$P_{virtual\ smear}(col,t) = P_{virtual\ smear}(col,t) - (dark\ level(t)) \cdot G_{virtual\ smear}(col,t) \cdot \left(\frac{t_{read}}{t_{exposure} + t_{read}} \right)$$

The available smear pixels for each column are tracked using the following logic:

Available Masked	Available Virtual	$C_{masked\ smear}$	$C_{virtual\ smear}$
True	True	1/2	1/2
True	False	1	0
False	True	0	1
False	False	0	0

$C_{masked\ smear}$ and $C_{virtual\ smear}$ are coefficients in the linear combination of the dark-corrected masked and virtual smear pixels (where ***G'*** are logical arrays with gaps = false):

$$C_{masked\ smear}(col,t) = \frac{1}{2} G_{masked\ smear}(col,t) \cdot (1 + G'_{virtual\ smear}(col,t))$$

$$C_{virtual\ smear}(col,t) = \frac{1}{2} G_{virtual\ smear}(col,t) \cdot (1 + G'_{masked\ smear}(col,t))$$

$$smear\ level(col,t) = P_{masked\ smear}(col,t) \cdot C_{masked\ smear}(col,t) + P_{virtual\ smear}(col,t) \cdot C_{virtual\ smear}(col,t)$$

The above smear and dark level estimates are computed during the collateral data calibration, resulting in a mean dark level value per channel and an array of smear levels per column per channel. These are later subtracted from the photometric pixels in each column:

$$P_{photometric}(col,t) = P_{photometric}(col,t) - (dark\ level(t)) - (smear\ level(col,t))$$

An additional complication to the smear level estimate is bleeding charge from saturated targets into the masked or virtual smear regions that are clearly visible in FFI data. CAL currently detects and gaps columns that are corrupted by bleeding charge in LC masked or virtual smear data (there are typically only one or two bleeding columns per channel).

3.7 Flat Field Correction

The flat field is the final major calibration step, and operates on photometric pixels to correct for spatial and temporal variations in pixel sensitivity to a uniform light source. Differences in pixel response can be due to variations in quantum efficiency or throughput changes in the field flattener lenses or anti-reflection coating of the CCD. The flat field model includes a geometric large-scale map combined with a small-scale (pixel-to-pixel) flat field map that is computed using a 9x9 pixel high pass filter[13]. The values represent the percent deviation from the local mean (with a median value across the focal plane[15] of ~0.96%), and the 2D flat field model is divided from the appropriate photometric pixels for each cadence:

$$P_{photometric}(row, col, t) = \frac{P_{photometric}(row, col, t)}{flat\ field(row, col, t)}$$

3.8 Additional Functionality in CAL

At the end of the last invocation of CAL for each CCD channel, the theoretical and achieved compression efficiency of the data are computed. These metrics, along with time series of black, smear, and dark level metrics also computed within CAL, are used by the photometer performance assessment (PPA)[8] module to track and trend data. The uncertainties can be computed within CAL by the propagation of uncertainties (POU) module[17]. The primary noise sources for *Kepler* include read noise, which is an additive noise source due to the readout process, quantization noise that is stochastic and results from quantization in the ADC and pixel requantization, and Poisson-like shot noise. The uncertainties in the raw pixel data are computed at the start of CAL, and (if enabled) POU runs in parallel with CAL and the uncertainties are propagated at each transformation step. If POU is disabled, the outputs to CAL are the raw uncertainties corrected only for gain.

4. SUMMARY AND FUTURE WORK

We have described the pixel-level corrections that are performed in the CAL Pipeline for LC, SC, and FFI flight data. The data corrections include: 2D and 1D black, gain, nonlinearity, undershoot and overshoot distortions from the LDE electronics, cosmic rays, bleeding charge, dark current, smear, and flat field variations. The algorithms were validated using simulated flight data from the End-To-End-Model[18] (ETEM) that was developed in the SOC, which simulates every layer of the data – from CCD and instrument artifacts to transit light curves – and has proven to be a powerful tool in the development and testing of the Pipeline modules. Output from CAL that is exported to the MAST includes raw and calibrated black, smear, and photometric pixels, along with the associated gap indicators and uncertainties. Additional metrics for cosmic ray detection, black level, smear level, and dark current level estimates are also provided.

Some issues that may be addressed in future versions of CAL include incorporating a more detailed time-varying 2D black model, refining the cosmic ray algorithms, and addressing bleeding charge in SC data. Due to the large volume of data that is processed each quarter (Q1, for example, yielded ~2TB for LC and SC CAL output alone), and the fact that the SOC will be reprocessing data throughout the duration of the mission, improving the performance of the algorithms is a high priority in order to maintain reasonable data processing times.

CAL has been successful at calibrating the flight data to high standards[15]. With data from just the first 10 days of flight, the *Kepler Mission* has detected and confirmed eight giant extrasolar planets to date[19] (including three that were known to exist in the *Kepler* field of view). The CAL algorithms and performance will continue to improve throughout the lifetime of the mission to support the search for smaller Earth-size planets.

ACKNOWLEDGEMENTS

The authors would like to thank Bill Borucki and Dave Koch for their decades-long dedication to *Kepler*, and the many others that contributed to the success of the mission. Special thanks to Sue Blumenberg for help with the preparation of the manuscript. Funding for the *Kepler* mission is provided by the NASA Science Mission Directorate.

REFERENCES

[1] Jenkins, J. M., *et al*., "Overview of the *Kepler* science processing pipeline," *ApJL* **713(2)**, L87–L91 (2010).
[2] Middour, C., *et al.*, "*Kepler* Science Operations Center architecture," *Proc. SPIE* **7740**, in press (2010).

[3] Twicken, J. D., *et al.*, "Photometric analysis in the *Kepler* Science Operations Center pipeline," *Proc. SPIE* **7740**, in press (2010).
[4] Twicken, J. D., *et al.*, "Presearch data conditioning in the *Kepler* Science Operations Center pipeline," *Proc. SPIE* **7740**, in press (2010).
[5] Jenkins, J. M., *et al.*, "Transiting planet search in the *Kepler* pipeline," *Proc. SPIE* **7740**, in press (2010).
[6] Wu, H., *et al*., "Data validation in the *Kepler* Science Operations Center pipeline," *Proc. SPIE* **7740**, in press (2010).
[7] Chandrasekaran, H., *et al.*, "Semi-weekly monitoring of the performance and attitude of *Kepler* using a sparse set of targets," *Proc. SPIE* **7740**, in press (2010).
[8] Li, J., *et al.*, "Photometer performance assessment in *Kepler* science data processing," *Proc. SPIE* **7740**, in press (2010)
[9] Bryson, S. T., *et al.*, "Selecting pixels for *Kepler* downlink," *Proc. SPIE* **7740**, in press (2010).
[10] Hall, J. R., *et al.*, "*Kepler* science operations processes, procedures, and tools," *Proc. SPIE* **7737**, in press (2010).
[11] Allen, C. A., *et al.*, "Focal plane characterization," *Proc. SPIE* **7740**, in press (2010).
[12] Haas, M., *et al.*, "*Kepler* science operations," *ApJL* **713(2)**, L115–L119 (2010).
[13] Van Cleve, J. and D.A. Caldwell, "*Kepler* Instrument Handbook"
[14] Akaike, H., "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, **AC-19 (6)**, 716–723 (1974)
[15] Caldwell, D. A., *et al.*, "Instrument performance in *Kepler's* first months," *ApJL* **713(2)**, L92–L96 (2010).
[16] Philbrick, R. H., "Correction of artifacts in correlated double-sampled CCD video resulting from insufficient bandwidth," *Proc. SPIE* **7244**, 72440M-72440M-12 (2009).
[17] Clarke, B. D., *et al.*, "A framework for propagation of uncertainties in the *Kepler* data analysis pipeline," *Proc. SPIE* **7740**, in press (2010)
[18] Bryson, S. T., *et al.*, "The *Kepler* end-to-end model: creating high-fidelity simulations to test *Kepler* ground processing," *Proc. SPIE* **7738**, in press (2010).
[19] Borucki, W., *et al.*, "*Kepler* planet-detection mission: introduction and first results," *Science* **327**, 977–980 (2010).