An Efficient End-To-End Model for the Kepler Photometer

Jon M. Jenkins,[a] Daniel J. Peters[b] and David W. Murphy[c]

[a]SETI Institute, M/S 244-30, NASA Ames Research Center, Moffett Field, CA 94035, United States;

[b]BATC, P.O. Box 1062 Boulder, CO 80306-1062, United States;

[c]MS 238-332, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, United States

ABSTRACT

The *Kepler* Mission is designed to characterize the frequency of Earth-sized planets in the habitable zones of solar-like stars in the solar galactic neighborhood by observing >100,000 main-sequence stars in a >100 square degree field of view (FOV) and seeking evidence of transiting planets. As part of the system engineering effort, we have developed an End-To-End Model (ETEM) of the photometer to better characterize the expected performance of the instrument and to guide us in making design trades. This model incorporates engineering information such as the point spread function, time histories of pointing offsets, operating temperature, quantization noise, the effects of shutterless readout, and read noise. Astrophysical parameters, such as a realistic distribution of stars vs. magnitude for the chosen FOV, zodiacal light, and cosmic ray events are also included. For a given set of design and operating parameters, ETEM generates pixel time series for all pixels of interest for a single CCD channel of the photometer. These time series are then processed to form light curves for the target stars and the impact of various noise sources on the combined differential photometric precision can be determined. This model is of particular value when investigating the effects of noise sources that cannot be easily subjected to direct analysis, such as residual pointing offsets, thermal drift or cosmic ray effects. This version of ETEM features extremely efficient computation times relative to the previous version while maintaining a high degree of fidelity with respect to the realism of the relevant phenomena.

Keywords: extrasolar planets, planet detection, transit photometry, performance modeling, Kepler Mission, space astronomy

1. INTRODUCTION

Scheduled for launch in 2007, the *Kepler* Mission is the 10th spaceprobe in NASA's Discovery Program and the first to seek out extrasolar planetary systems. Consisting of a 0.95-m, Schmidt-type telescope and a focal plane of 42 CCDs, *Kepler* will observe ≥100,000 stars in a >100 square degree field of view centered on 70° galactic longitude, +5° galactic latitude in the constellation of Cygnus. The principal goal of this mission is to identify the signatures of transiting planets in the habitable zones of their parent stars. The fractional drop in stellar brightness due to a transiting planet is simply the ratio of the disk area of the planet to the disk area of its star. For an Earth-sized planet transiting a Sun-like star, this ratio is 1 part in 12,000. The probability that the planetary orbital plane will permit an arbitrarily placed observer to view the transits is the ratio of the stellar radius to the mean orbital distance of the planet. For a planet orbiting a Sun-sized star in a 1-AU orbit, this probability is ∼0.5%. Taken together, these two facts imply that *Kepler* must observe thousands of main-sequence, late-type stars with 20 ppm precision on timescales of hours to catch its quarry. The point design for *Kepler* is to achieve a 4 σ signal to noise ratio (S/N) for a single transit of an Earth-sized planet transiting a m_R=12 G2V star. For four transits the total S/N is 8 σ, yielding a detection rate of 84% for a detection threshold of 7 σ, which controls the total number of expected false alarms to no more than one for the entire experiment. The large number of target stars will ensure that *Kepler* returns significant results even in the event that no Earth-sized, habitable planets are detected. The development of the *Kepler* Mission is well underway with the System Requirements Review completed in October

Further author information: (Send correspondence to J.M.J.)
J.M.J.: E-mail: jjenkins@mail.arc.nasa.gov, Telephone: 1 650 604 1111

2003, and the Preliminary Design Review and the Critical Design Review scheduled for October 2004, and June 2005, respectively. As part of the development effort, three analytical tools have been constructed to aid in the design process.

The first tool is the Combined Differential Photometric Precision (CDPP) spreadsheet, which tracks the CDPP for a 6.5-hour transit for a G2V star as a function of the apparent stellar magnitude and the set of mission design parameters. The quantity CDPP is the effective white noise standard deviation in a 6.5-hour interval that determines the S/N of a 6.5-hour transit of a given depth. For example, a CDPP of 20 ppm for a star with a planet exhibiting 84 ppm transits lasting 6.5 hours leads to a single transit S/N of 4.1 σ. The second tool is the *Kepler* Merit Function, which assesses the value of the science return in terms of the number of expected planetary detections for an assumed planetary population distribution (including both habitable and non-habitable planets) and in terms of the number of stars for which pressure mode (p-mode) oscillations can be studied. The third tool is the End-To-End Model, which is a Monte Carlo simulation of the *Kepler* Photometer producing synthetic data time series at the pixel level. These three tools are supported by a Noise and Error List that tracks the effects and magnitudes of $\sim$75 separate stochastic and systematic noise sources that potentially affect *Kepler* photometric performance. The first two tools do not produce time series but rather perform analyses combining the expected noise and signal properties or predicting the science return given these properties. As such, ETEM is extremely valuable in assessing the effects of noise terms which are not subject to closed form analysis, such as the effects of pointing "jitter" or the effects of residual cosmic ray events on the CDPP. This suite of complementary analytical tools allows us to predict the performance of *Kepler* and to measure the effect of design choices on the scientific return.

The goal of this paper is to describe the methodology behind the current version of ETEM and to give examples of the investigations it has enabled. ETEM has been under development since 1995 as a collaboration between *Kepler* team members at both NASA Ames Research Center and at Ball Aerospace & Technologies Corporation (BATC). ETEM began as a FORTRAN program that generated synthetic photometric data for a single star at the pixel level given a Point Spread Function (PSF), characterization of observational noise sources such as sky background, read noise, and dark current, and a sample pointing offset time series. This simple model of an isolated star was later incorporated into a larger program to simulate 100 target stars together with the effects of shutterless operation as part of the *Kepler* Technology Demonstration (KTD) funded by the NASA Discovery Program.[1,2] Generating synthetic photometric data is a necessary component of assessing the expected photometric performance. The back end of ETEM consists of analyzing the output data to construct light curves and measure the photometric precision achieved for the input design. This analysis uses algorithms similar to those baselined for the actual data processing and analysis system. These differences exist because the output of ETEM has been until recently restricted to producing rather short data sets of up to a few days for relatively few stars compared to the actual mission data. Over the past two years we have redesigned ETEM in such a way as to permit the modeling of millions of background stars together with up to 2000 target stars on a single CCD readout area for artificial data runs of up to three months. In addition to its significantly improved efficiency, the current ETEM incorporates a great deal more realism accumulated as the spacecraft and photometer design have matured.

Figure 1 depicts the logical flow of ETEM beginning with the steps required to set up the model for the given system design parameters and ending with the generation of synthetic photometer data. The top line of the logic flow contains the sequence of steps necessary to prepare ETEM for generating synthetic photometer data. These steps incorporate all the phenomena relevant to the quantification of the photoelectronic image read out from a CCD prior to digitization by the Analog to Digital Converter (ADC), and the addition of stochastic noise. The second row of the flow consists of exercising the model to produce synthetic images, add various random noise sources such as shot noise and read noise, digitize the result, and finally, write the results to disk. This sequence of steps is repeated until the requested data set is completed.

The major improvements to ETEM reported in this paper lie in the numerical approach taken to model the response of the CCD pixels to motion of the stellar images. The perturbations to the image position due to tracking errors of the Attitude Determination and Control System (ADCS) and to astrophysical effects are either small enough to be well modeled by polynomial expansions of the pixel values in terms of the (Δx, Δy) offsets, or occur on timescales much longer than that of a transit duration, and hence, are unimportant with respect to determining S/N of a transit. The principal purpose for ETEM is to analyze noise sources impacting the detectability of transits, although future enhancements envisioned include modeling of long term astrophysical effects such as differential velocity aberration to better understand the complications these might pose with respect to the data processing and target management.

In order to achieve this efficiency, some simplifying assumptions were made. The stars are assumed to move together, that is, no provision is made for parallax, proper motion, or second order effects of differential velocity aberration. As argued above, these effects do not significantly affect the S/N of a transit, but we do need to understand how they might effect

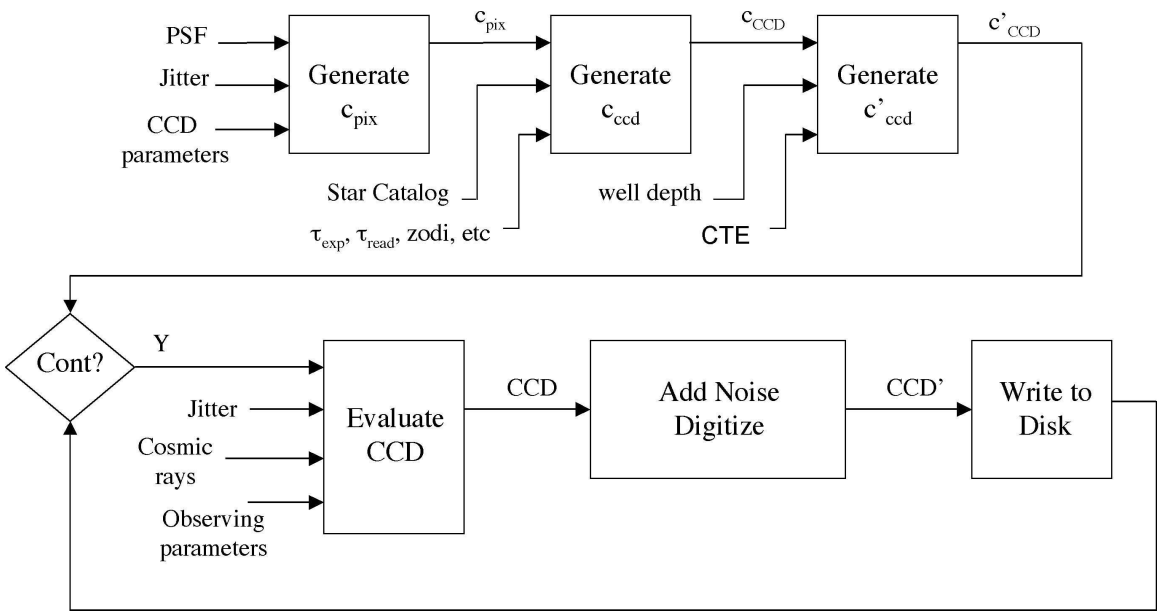

Figure 1. Logical flow chart for ETEM. The top flow consists of the sequence of steps necessary to set up ETEM to generate synthetic images, incorporating the relevant design parameters such as PSF, CCD dimensions, pixel-to-pixel sensitivity, intrapixel sensitivity, integration time, readout time, etc. The bottom flow consists of generating synthetic noise-free images, adding stochastic noise, digitizing the result and writing the data to disk. This is repeated until the requested data set is completed.

operations and analysis of the data. Intrinsic stellar variability is also not currently modeled for the target or background stars. Stellar variability for old, main-sequence stars like the Sun occurs on timescales much longer than the duration of a transit of a terrestrial planet. The results from ETEM and the CDPP spreadsheet have been incorporated into detailed studies of the effects of solar-like variability on the detectability of transiting Earth-like planets.[3] Provisions have been made in ETEM to allow for the inclusion of stellar variability into the simulations. Doing so would necessarily reduce the efficiency of ETEM. It is unclear whether this is necessary. To date, ETEM is used mainly to determine the contribution of noise sources not amenable to analysis. As such, once a noise term is characterized by ETEM, its effects are then incorporated into the CDPP spreadsheet and into the Merit Function. Although we plan to evolve ETEM to incorporate more realism over time, it is likely that *Kepler* will continue to require and exercise several numerical models during development.

This paper is organized as follows. Section 2 describes the steps taken to develop polynomial representations for a CCD readout area. Section 3 details the generation of synthetic CCD data including the addition of stochastic noise to the frames. Section 4 summarizes the results and discusses future work.

2. A POLYNOMIAL REPRESENTATION FOR THE *KEPLER* PHOTOMETER

This section details the phenomena incorporated into the first two boxes labeled "Generate $\mathbf{c}_{CCD}$" and "Generate $\mathbf{c}'_{CCD}$" in Fig. 1. With the exception of the effect of spilling of saturated charge, all the phenomena modeled in ETEM for generating synthetic CCD images are linear, so that they can be directly incorporated into a polynomial representation for the response of a CCD to image motion. Nevertheless, most pixels' behavior is well modeled by a polynomial representation, and those few pixels that are not, can be handled separately.

2.1. Response of Pixels to Image Motion

As in previous versions of ETEM, the first step is to determine how the pixels under a stellar image respond to image motion. For a star of brightness I_0 located at (x_0, y_0), the charge that is developed on the CCD is

$$I_\lambda = [I_0 PSF_\lambda (x - x_0, y - y_0)\, S_\lambda (x, y)] * D_\lambda (x, y), \tag{1}$$

where λ is the wavelength, $PSF_\lambda(x,y)$ is the PSF, $S_\lambda(x,y)$ is the sensitivity function of the CCD, $D_\lambda(x,y)$ is the diffusion kernel, and '$*$' denotes the convolution operator. The optical PSF is derived from the optical design of the photometer using a raytracing algorithm (ASAP) for each of 21 wavelengths across the *Kepler* bandpass (420–860 nm).[4]

For $S_\lambda(x,y)$, we take the results reported by Ref. 5, which are only reported at two wavelengths, 600 and 850 nm. Below 600 nm, the variation of S_λ is quite small, while at 850 nm, the peak-peak variation is $\sim$10%. Fortunately, models for $D_\lambda(x,y)$, developed at BATC for the *Kepler* flight CCDs indicate that diffusion is important only for wavelengths shortward of 600 nm, and is not apparent at longer wavelengths.[6] This is due to the fact that the longer wavelength light travels through the entire CCD thickness and is absorbed in or very near the active region, so that there is little opportunity for diffusion from the absorption site before readout. At these wavelengths, S_λ exhibits variations that are consistent with the physical gate structure of the CCD. Conversely, this also explains the relative unimportance of intrapixel sensitivity variations for short-wavelength light, which is absorbed above the active region and must diffuse down into it prior to readout. The "blue" light never sees the gate structure, which has the opportunity to scatter the "red" light.

Note that the apparent complementarity of D_λ and S_λ implies that I_λ can be expressed as a cascade of convolutions involving PSF_λ, S_λ restricted to a single pixel, and either D_λ or S_λ. Once I_λ is determined at all 21 wavelengths, it can be weighted by the stellar spectrum and photometer bandpass response and summed over λ to determine the total charge intensity for each pixel as a function of position. In the current version of ETEM, we have modeled the process using a total optical PSF over the solar spectrum and photometer bandpass, and an effective charge diffusion kernel prior to consideration of the CCD pixel sensitivity. We have performed analyses to show that this approach is conservative, but are working to improve the fidelity of this step as per the discussion above.

In previous versions of ETEM, the importance of charge diffusion was not recongnized and hence, was ignored. This is actually a conservative assumption in that the charge diffusion blurs the optical PSF and reduces somewhat the sensitivity of the pixel values to motion. The values of each pixel as a function of $(\Delta x, \Delta y)$ offsets from a nominal position were determined by scaling a tabular representation of the optical PSF (on a 5 by 5 pixel region) with 13 by 13 subpixel resolution, scaling it by the intrapixel sensitivity and then integrating over each pixel region. This response was evaluated at a particular image offset for a given jitter time series by bilinear interpolation over the tabulated values. The process of interpolation is numerically quite intensive especially when using cubic or spline interpolation. We note that this process yielded small but not insignificant modeling errors as the bilinear interpolation actually used did not preserve flux for a perfectly uniform CCD response. The most significant improvement in computational efficiency for ETEM lies in recognizing that for *Kepler*, the expected perturbations to the CCD images due to pointing "jitter", thermal drifts and astrophysical effects such as differential velocity aberration are quite small over timescales of seconds to several days. For example, the pointing offset "jitter ball" is required to be no larger than 0.1 arcsec (or 2.5 mpix), 3 σ, and is expected to be much smaller in practice. For such small pointing offsets, the response of the pixels to image motion is smooth and well represented by low-order, two-dimensional polynomials.

The current version of ETEM takes advantage of this fact and incorporates a polynomial fit to the response of each pixel to motion of a stellar image over a fine grid containing the "jitter ball", resulting in pixel polynomial coefficients, c_{pix}. For any given pointing offset within the design region, each pixel value can then be determined by evaluating the corresponding polynomial for a given pointing offset pair $(\Delta x, \Delta y)$ and simply scaling the result to an intensity appropriate for a given magnitude star. Figure 2 shows the fitting error between the polynomial representation and the cubic-spline interpolation of the pixel response for the required jitter Power Spectral Density (PSD). For the required Attitude Determination and Control System (ADCS) performance, we find that 3rd order polynomials adequately represent the pixel response to motion. The polynomial is of the following form

$$p(\Delta x, \Delta y) = c_{00} + c_{10}\Delta x + c_{01}\Delta y + c_{20}\Delta x^2 + c_{11}\Delta x\Delta y + c_{10}\Delta y^2 + c_{30}\Delta x^3 + c_{21}\Delta x^2\Delta y + c_{12}\Delta x\Delta y^2 + c_{03}\Delta y^3. \quad (2)$$

Higher order polynomials can be applied to provide better fits or to allow for a larger design range of jitter. Once the polynomials are determined, an entire CCD frame can be populated with stars using a realistic stellar distribution.

2.2. Stellar Population of a CCD Frame

A synthetic star catalog is used to populate a single CCD readout channel consisting of 1100 columns by 1024 rows.* The polynomials for each pixel in a CCD can be determined by simply adding together the pixel polynomials for all the stars

*The actual flight CCDs have 2200 columns and 1044 rows with dual readout amplifiers. The bottom 20 rows are masked to allow for estimation of and correction for the effects of shutterless readout.

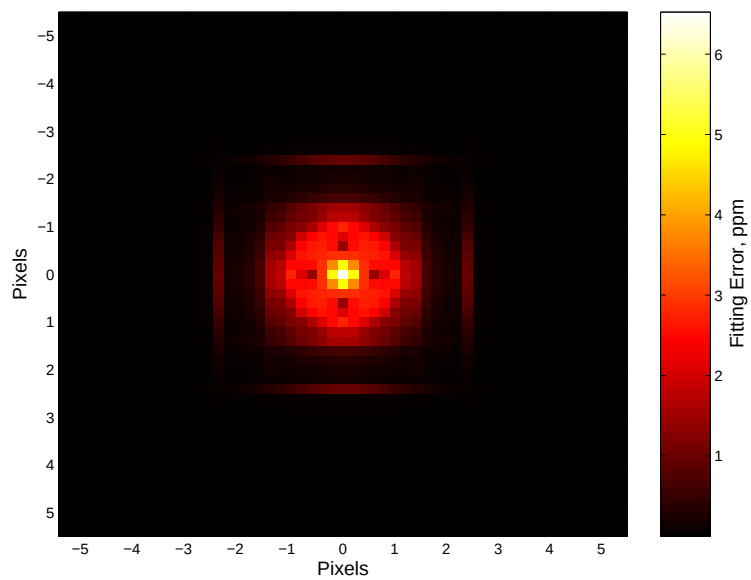

Figure 2. The rms fitting error between a cubic spline interpolated representation of response of a pixel to image motion of a stellar PSF, and a 3rd order polynomial representation. The greatest errors occur near the PSF core but at the 10^{-5} level are not significant.

whose images fall on a given pixel.

Following Ref. 7, we make use of galactic models made publicly available by the Observatoire de Besançon[†] (see, e. g., Refs. 8, 9, and 10) to obtain expected star counts as a function of apparent magnitude, spectral type and age. The USNO-A2.0 database yields 223,000 stars to m_R=14.0 in the 106 square degrees of *Kepler's* FOV.[11] This establishes an appropriate mean extinction of ∼1.0 mag kpc^{-1} for the Besançon model. We note, however, that the bandpass for *Kepler* extends from ∼0.45 to ∼0.85 μm, which is far wider than the bandpasses available for the Besançon models. For the purpose of counting stars, using the R band should reflect the number of stars of greatest interest, but may tend to undercount the number of late main sequence stars. Figure 3 shows the distribution of stars of all luminosity classes and spectral types predicted by the Besançon model for *Kepler's* FOV.

To construct the CCD polynomial c_{ccd}, we first generate c_{pix} for 25 different nominal centerings of stars within their central pixel, on a 5 by 5 subpixel grid. A synthetic star catalog is compiled by sampling the distribution provided by the Besançon model, drawing random coordinates for each star in the CCD's FOV, and partitioning the stars into 25 polynomial classes. For each polynomial class, each CCD coefficient frame is determined by adding the stellar intensity to the center pixel on a blank 1100 by 1024 pixel array, and then convolving this "impulse frame" with each 11 by 11 coefficient array for each of the 10 polynomial coefficient planes (assuming 3rd order polynomials). In this way, efficient Fast Fourier Transform (FFT) methods can be used to assemble each coefficient frame for each of the 25 stellar polynomial classes. The resulting polynomial frames are added together sequentially for each polynomial class as they are computed.

Given a pointing offset matrix A_{jit} and the CCD coefficients, $\mathbf{c}_{CCD}$[‡], the charge deposited in a given interval of time is $CCD = A_{jit}\,\mathbf{c}_{CCD}$. In ETEM, $\mathbf{c}_{CCD}$ is scaled so that evaluation of the polynomial yields flux in e$^-$ s^{-1}. Figure 4 shows a 2-D histogram of a realization of a jitter time series for the expected ADCS performance binned to 2 Hz sampling. We are now in a position to generate synthetic CCD images for a pointing offset time series.

[†] http://www.obs.-besancon.fr/modele/modele.ang.html

[‡] Note that the polynomial coefficients are denoted by boldface. This is to indicate that the $\mathbf{c}_{CCD}$ is a matrix whose columns correspond to the polynomial coefficients, and whose rows correspond to each of the pixels under analysis. The evaluation of the CCD polynomial can then be expressed using matrix algebra, although the results must be reshaped to recover the original dimensions of the CCD.

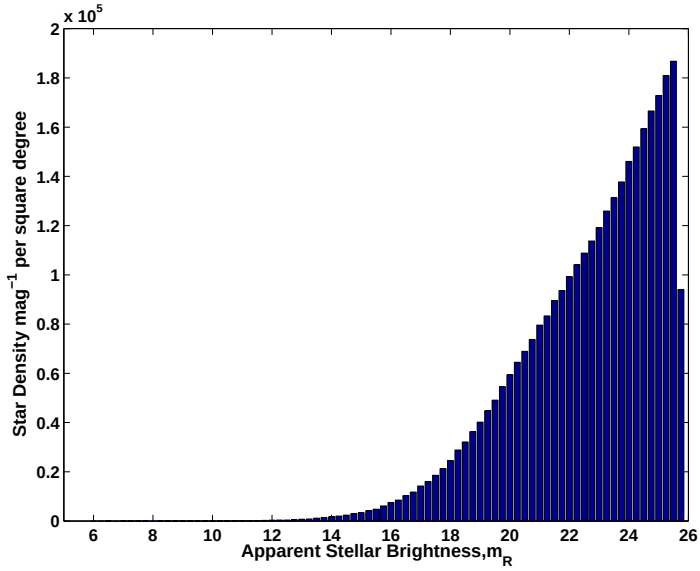

Figure 3. A histogram showing the distribution of the density of stars with apparent R magnitude. All luminosity classes and spectral types are represented.

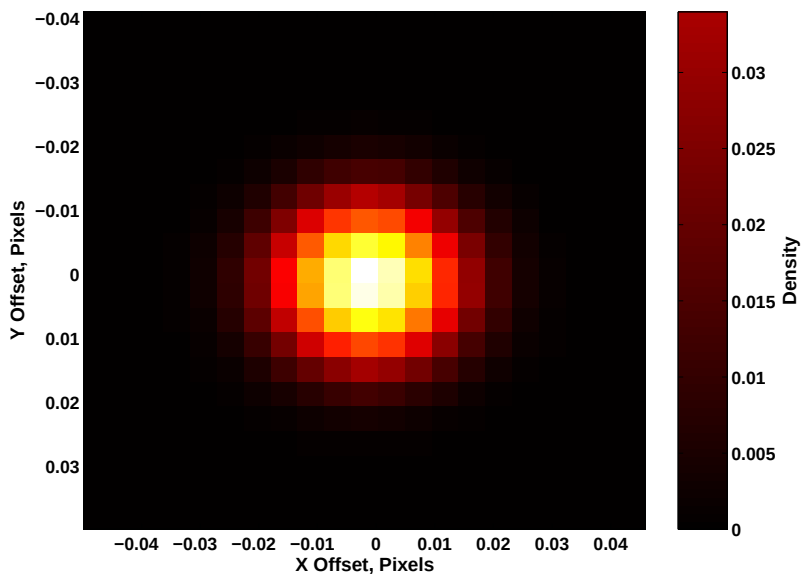

Figure 4. A 2-D histogram showing the distribution of pointing offsets for the expected ADCS performance for *Kepler* binned to 2 Hz sampling. The standard deviation of the pointing is 0.01 pix in each axis.

2.3. Additional Imaging Phenomena

The polynomial representation $\mathbf{c}_{CCD}$ developed in the previous section allows us to evaluate the charge developed on a CCD for a given attitude, but it does not factor in all the relevant effects. In particular, we need to accommodate additive noise sources such as dark current, zodiacal light and the effect of the shutterless operation. Saturation effects must also be considered, along with Charge Transfer Efficiency (CTE), but these are the subjects of Sect. 2.4.

There are two purely additive fluxes that do not respond significantly to image motion: dark current and zodiacal light. Dark current accumulates during exposure and readout of the CCDs and is a strong function of the operating temperature of

the CCD. Although the operating temperature of *Kepler's* focal plane is so cold ($< -90^\circ$C) that the dark current is expected to be negligible, it is still accommodated in ETEM. The zodiacal light is solar flux that is scattered from dust grains in and above the ecliptic plane into the Photometer's aperture. Characterization of zodiacal light by the Hubble Space Telescope implies that the zodiacal background will inject the equivalent of an m_R=19 star in every CCD pixel (4 arcsec by 4 arcsec). This is much higher than the expected dark current. Neither dark current nor zodiacal light will vary with the expected pointing errors, although zodiacal light will vary smoothly over large spatial scales and on time scales of months. For the time scales of most interest to ETEM, these flux sources can be simply added to the constant term in $\mathbf{c}_{CCD}$.

At this point ETEM incorporates pixel to pixel sensitivity variations. This is particularly simple as it amounts to scaling each pixel polynomial by the relative sensitivity of the pixel. Most ETEM runs use a highly conservative value of 5% for the interpixel sensitivity variations and draw each pixel's relative sensitivity from a Gaussian distribution.

The fact that *Kepler* lacks a shutter has significant but mostly benign implications for the CCD images. During readout, each row is clocked down the CCD, passing under any stars falling on the CCD below their position during the exposure. At the same time, new rows are being read in from the top of the CCD and clocked down to their nominal locations for the next exposure, passing underneath stars above their exposure positions. The resulting images contain vertical streaks due to star light accumulating in the pixels along each column during readout. The smear component can be calculated by summing each frame of coefficients along the columns, scaling for the exposure time spent in each row. The smear polynomial only responds to image motion along the rows, except at the very edges of the CCD. Accounting for smear in $\mathbf{c}_{CCD}$ amounts to replicating the row polynomial for smear and adding it to each of the rows in $\mathbf{c}_{CCD}$. There is a provision for overclocking the CCDs by 20 rows for testing purposes, but also to allow for a separate estimate of smear. Such overclocked rows do not exist during the exposure, so while they pick up smear as they are clocked through the field, they do accumulate some dark current during readout.

Another source of flux exists: scattered light in the photometer. Studies have been performed to estimate the fraction of the focal plane that will be adversely affected by ghost images from the handful of $m_R \leq 6$ stars in the FOV. At this point, however, the design is not mature enough to quantify and model the effect in ETEM. The methodology used to model the star field applies to the ghosts and will be used to model the effect once sufficient data is collected during testing.

2.4. Saturation and CTE

All the previous phenomena represented linear transformations of the polynomials representing the response of the CCDs to image motion. At this point it is necessary to include the nonlinear effect of charge saturation and CTE in the simulation. We model saturation of a pixel as a process that conserves charge, but distributes it along the column containing the saturated pixel evenly in both directions. The former effect is supported by experiments performed with the Kepler Tech Demo and with HST.[12] For the present purposes, it is not important to have a model for saturation that is realistic in all details. It is sufficient to have a model that is indicative of the difficulties pixel saturation may pose. Saturation will only affect a small handful of target stars in any event.

In ETEM, after the effects described in Sect. 2 are accounted for, a set of images is generated over a grid of offsets, much as for the calculation of the original pixel polynomials. For these images, pixels that exceed the specified CCD well depth are iteratively spilled up and down their columns until no pixels are saturated. The imperfect CTE is modeled at this point by noting that it can be expressed as a linear infinite impulse response (IIR) digital filter. Let b'_n be the pixel value read out from the CDD including the effects of CTE, and let $\{b_n, b_{n+1}, b_{n+2}, \ldots\}$ be the pixel values in sequence of readout starting with pixel n before including the effects of CTE. We can express b'_n in terms of the $\{b_n, b_{n+1}, b_{n+2}, \ldots\}$ as

$$b'_n = \alpha\, b_n + (1-\alpha)\, b_{n+1} + (1-\alpha)^2\, b_{n+1} + \ldots, \tag{3}$$

where α is the fraction of charge in a pixel that is successfully clocked to the next row for a single clock cycle. Although the effective CTE filter is IIR, a typical value for α is 0.9996, so that $(1-\alpha)^m$ becomes insignificant for $m > 8$. The CTE filter is convolved with each column of the images for the parallel readout and with each row for the serial readout.

New CCD pixel polynomials are fitted to the set of images and the fitting residuals are examined for poorly behaved residuals. Saturated pixels and neighboring pixels that accept spilled charge are typically flagged, and the spill of saturated charge and CTE are modeled directly for these pixels and their neighbors. All other pixels' behavior is well represented by the new polynomials, $\mathbf{c}'_{CCD}$, since all the transformations, including the effect of CTE, are linear transformations of the original polynomials. At this point, ETEM is ready to generate synthetic photometric data for a specified run.

3. RUNNING ETEM

This section describes the steps performed to generate synthetic CCD data once the development of the polynomial representation for the CCD response to motion is complete. To generate synthetic CCD data, ETEM evaluates the polynomial $\mathbf{c}'_{CCD}$, simulating spill of saturated charge and CTE for flagged pixels. Shot noise and read noise are added to the pixels, along with charge from cosmic ray events, if desired. The results are digitized, and are then written to disk and the process is repeated until the run is complete. There are two modes of operation for ETEM with respect to the generation of synthetic data, and these relate to operational constraints for *Kepler*.

To prevent saturation of target stars, the exposure time for the photometer is ∼3 s, so that each day approximately 29,000 images are acquired. There is not enough memory on the Solid State Recorder (SSR) onboard *Kepler* to keep all this data, so two lossy compression techniques are used to reduce the size of the data set. The first technique is to co-add the images for 15 minutes, reducing the total number of images stored on the SSR per day to 96. For the second technique only the pixels of interest are stored: those containing target stars and collateral pixels used to correct for CCD artifacts and other systematic errors, such as sky background, dark current and smear from shutterless operation. So, too, for ETEM there is no reason to generate data for pixels that won't be analyzed later. An analysis module examines the pixel content for each target star and determines the optimal photometric aperture in a similar manner to that described in Ref. 1.

The two modes relate to the generation of 15-minute frames, or long cadences, onboard *Kepler*. In the first mode, individual readouts are generated explicitly by evaluating $\mathbf{c}'_{CCD}$, adding stochastic noise and digitizing the results. These are co-added until the appropriate number have been summed to form a long cadence, then the results are written to disk. This mode is useful in examining phenomena that operate on timescales shorter than 15 minutes, such as analyzing the ability of *Kepler* to identify and deal with cosmic rays. The other mode of operation is to evaluate $\mathbf{c}'_{CCD}$ for an entire 15-minute interval, called the long cadence mode, add all the stochastic noise corresponding to that interval, and to model the effects of quantization by adding additional random deviates which are drawn from an appropriate distribution. For long runs of ETEM, the long cadence mode is preferred as it is ∼300 times less computationally intensive than the first.

The long cadence mode is enabled by the polynomial representation itself and the fact that the noise on a 15-minute frame can be analytically related to that at the single exposure level. Consider the process of co-adding a sequence of noise-free CCD images generated by evaluating $\mathbf{c}'_{CCD}$. Let $\mathbf{b}_n$ be a sequence of noise-free CCD frames constructed by evaluating polynomial $\mathbf{c}'_{CCD}$. For example, suppose we wish to bin the results by a factor of three, yielding $\tilde{\mathbf{b}}_\mathbf{n}$. This process can be written as

$$\tilde{\mathbf{b}} = BA_{jit}\,\mathbf{c}'_{CCD}, \tag{4}$$

where B implements the binning operation and is given by

$$B = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & \dots & 0 & 0 & 0 \\ & & & \vdots & & & \ddots & & \vdots & \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 \end{bmatrix}. \tag{5}$$

Note that the associative property of matrix multiplication implies that A_{jit} can be pre-multiplied by the binning matrix B before it is multiplied by $\mathbf{c}'_{CCD}$. A significant reduction in processing time for the 15-minute mode is achieved relative to the single exposure mode by forming $\tilde{A}_{jit} = BA_{jit}$, and hence, binning the jitter matrix prior to evaluating $\mathbf{c}'_{CCD}$.

Whether or not ETEM is operating in the single exposure mode, once $\mathbf{c}'_{CCD}$ is evaluated, it is time to add shot noise and read noise. This is accomplished by adding Gaussian noise of appropriate variance to the noise-free polynomial values. In the long cadence mode, the specified single-exposure read noise is scaled by the square root of the number of exposures in a 15-minute integration. At this point, synthetic cosmic rays can be added to the images, if desired.

3.1. Cosmic Ray Events

The cosmic ray flux environment has been of great concern to almost all space missions with CCDs since they are sensitive to cosmic rays. The actual flux experienced by a device depends a great deal on the exact orbit, that is, is the spacecraft in low Earth orbit (LEO), or is it in deep space? The flux also depends on the shielding and configuration of the detectors within the spacecraft, which can affect the generation of secondaries from primary events. In any case, the *Kepler* Mission has adopted a flux rate of 5 $\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$ based on previously flown missions in similar orbits, such as *SOHO*.

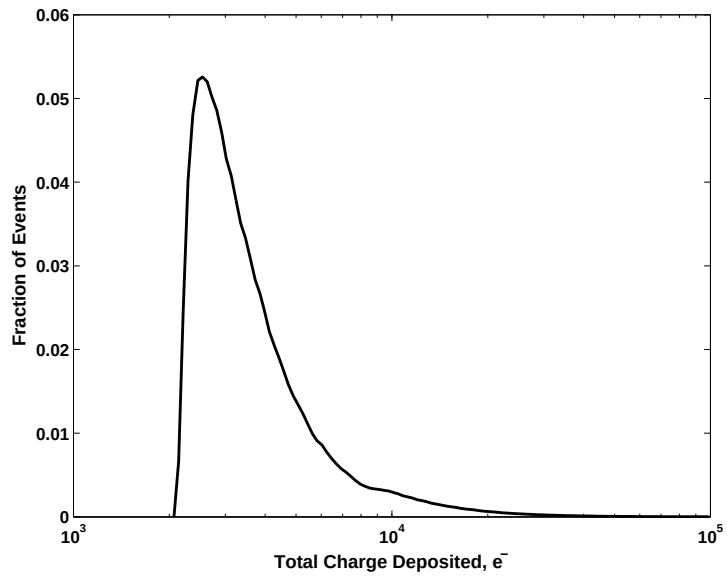

Figure 5. The distribution of the total charge deposited into a *Kepler* CCD per cosmic ray event. This distribution resulted from modeling taking into account the expected cosmic ray environment for *Kepler's* orbit, and a detailed structural model for the spacecraft, the instrument, and any planned radiation shielding.[13]

A study was conducted by BATC for *Kepler* to derive the distribution of total charge deposited into a CCD for each cosmic ray hit.[13] The results are displayed in Fig. 5, which shows a mode of $\sim$2500 e$^-$, little or no charge below $\sim$2000 e$^-$, and a long upper tail trailing out to at least 100,000 e$^-$. We note that 90% of events deposit less than 6,200 e$^-$ into a CCD. To put this into perspective, note that an m_R=12 star occupies about 25 pixels, and that over 6.5 hours, about 4×10^9 e$^-$ accumulates in its aperture. The shot noise for such a star will be 63,245 e$^-$. Now, 25 pixels receive a cosmic ray flux rate of 21.3 per 6.5 hr interval. To compare this to the effect of uncorrected cosmic rays, we need to model the distribution of charge from cosmic rays at the pixel level.

To transform the distribution in Figure 5 to the desired one, we applied the following assumptions: 1) The total charge deposited is uniformly distributed over the path traveled by the cosmic ray as it traverses the CCD slab. 2) The charge desposited by the cosmic ray diffuses the same way as does charge from actual photons. Given the geometry of the CCDs (27 μm $\times$ 27 μm $\times$ 16 μm), we traced random rays through a 13 by 13 pixel region of a CCD, distributing the charge in each pixel encountered by a ray according to the assumptions above. The CCD pixels were divided into 13×13 sub-pixels for the purposes of the numerical calculations. We amassed a catalog of 6,097 cosmic ray trails, normalized so that the sum of each trail was unity. Each trail, then, can be scaled by a random deviate drawn from the total charge distribution to model the effect of a single cosmic ray. A Monte Carlo experiment using this model showed that the rms noise injected by cosmic rays in a 25-pixel aperture in a 6.5-hr interval is 21,171 e$^-$, or about 5 ppm relative to the stellar flux. This is not significant compared to the shot noise.

3.2. Digitization of the Synthetic, Noisy CCD Frames

After the stochastic noise has been added to the synthetic CCD frame, it can be digitized and either co-added to the running sum, or written to disk, in the long cadence mode. For this latter mode, the effect of quantization at the single exposure level can be modeled by adding zero-mean, White Gaussian Noise (WGN) with a standard deviation equal to $\sqrt{M}G/\sqrt{12}$ where G is the gain in e$^-ADU^{-1}$, and M is the number of co-adds. This does not accurately model extremely dim pixels whose exposure-to-exposure variations are less than 1 ADU, but these do not occur in target star pixels. In this mode, the final step is to normalize the pixel values by the gain to convert the scale to ADU from e$^-$. In the single exposure mode, the digitization can be performed explicitly. Note that for the single exposure mode, there is the opportunity to include the effects of nonlinearities in the analog signal processing chain before the quantization.

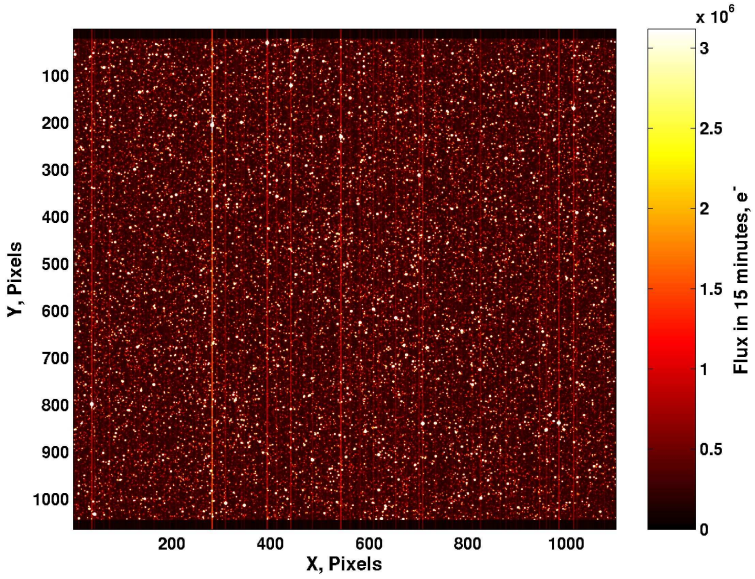

Figure 6. Synthetic accumulated CCD frame for *Kepler*. The image is the mean 15-minute frame for a synthetic stellar population generated by ETEM, clipped to 1% of the full range. Approximately 1×10^6 stars are simulated.

Figure 6 displays the average 15-minute frame for one run of ETEM, while Fig. 7 displays a single, 2.88 s exposure where only the pixels of interest have been calculated. The effects of the shutterless readout are evident as vertical streaks. In ETEM, the long cadence pixels of interest are written to disk and then subjected to analysis to determine the CDPP. By comparing the results of separate runs with individual noise sources toggled on and then off, it is possible to assess their contribution to the total CDPP budget.

4. CONCLUSIONS

We have presented the mathematical approach taken for the current version of ETEM. The use of a polynomial representation for the response of the CCD pixels to image motion allows for almost all the salient operational and astronomical effects to be incorporated directly into the polynomial representation. This is due in part to the precise pointing stability afforded by the benign, Earth-trailing, heliocentric orbit, and the simple, fixed attitude. This polynomial representation leads to an efficient simulator with the ability to include the effects of millions of background stars on the chosen target stars. The efficiency of ETEM has been improved by a factor of greater than 100, allowing for much longer runs to be computed for many more target stars with significantly improved realism. This version of ETEM has been used to study a number of important issues affecting the design of the *Kepler* Photometer, including the effect of the PSF on the CDPP, the effect of cosmic rays on CDPP, the ability of the baseline cosmic ray detection and correction algorithm to deal with cosmic rays at the single exposure level, and the sensitivity of the CDPP to nonlinearities in the analog processing chain and in the Analog to Digital Converters (ADC's). Future efforts include the generation of sample data for testing the flow of data and data processing in the photometer and through the science processing pipeline of the ground segment, including the Data Management Center at the Space Telescope Science Institute and the Science Operations Center at NASA Ames Research Center. While we plan to further improve ETEM to incorporate effects such as differential velocity aberration, it is of sufficient fidelity to address important systems engineering questions arising at this stage of the development cycle.

ACKNOWLEDGMENTS

We thank Rob Philbrick, Neal Nickles, David Koch, and Ron Gilliland for their valuable input on matters discussed in this paper. Funding for the *Kepler* mission is from the NASA Discovery program UPN 853-15-01.

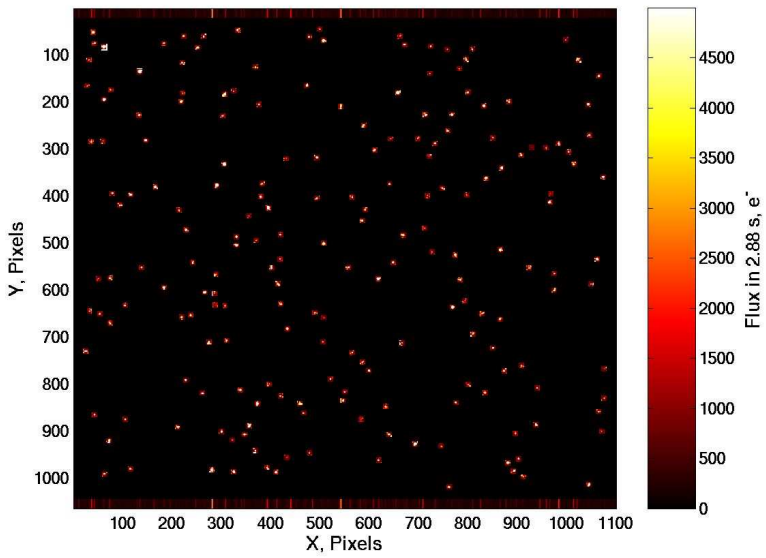

Figure 7. An image representing a single 2.88-s exposure generated during one run of ETEM, clipped to 0.5% of the full range. For effi ciency, only pixels for those stars selected for study are calculated during the run, along with collateral pixels allowing for estimation and removal of dark current and shutterless smear.

REFERENCES

1. J. M. Jenkins, F. Witteborn, D. G. Koch, E. W. Dunham, W. J. Borucki, T. F. Updike, M. A. Skinner, and S. P. Jordan, "Processing CCD images to detect transits of Earth-sized planets: maximizing sensitivity while achieving reasonable downlink requirements," in *UV, Optical, and IR Space Telescopes and Instruments*, J. B. Breckinridge, and P. Jakobsen, eds., *Proc. SPIE* **4013**, pp. 520–531, 2000.
2. Q. P. Remund, S. P. Jordan, T. F. Updike, J. M. Jenkins, and W. J. Borucki, "Kepler System Numerical Model for the Detection of Extrasolar Terrestrial Planets," in *Instruments, Methods, and Missions for Astrobiology IV*, R. B. Hoover, G. V. Levin, R. R. Paepe, and A. Y. Rozanov, eds., *Proc. SPIE* **4495**, pp. 182–191, 2002.
3. J. M. Jenkins,"The Impact of Solar-like Variability on the Detectability of Transiting Terrestrial Planets," *Astroph. J.* **575**, pp. 493-505, 2002.
4. D. Koch, W. Borucki, E. Dunham, J. Geary, R. Gilliland, J. Jenkins, D. Latham, E. Bachtell, D. Berry, W. Deininger, R. Duren, N. Gautier, L. Gillis, D. Mayer, C. Miller, D. Shafer, C. Sobeck, C. Stewart, M. Weiss, "Overview and status of the Kepler Mission", in *Proc. SPIE* **5497**, 2004.
5. P. R. Jorden, J.-M. Deltorn, and A. P. Oates, "Nonuniformity of CCDs and the effects of spatial undersampling," in *Instrumentation in Astronomy VIII*, D. L. Crawford, and E. R. Craine, eds., *Proc. SPIE* **2198**, pp. 836–850, 1994.
6. Personal communication, Rob Philbrick, 2004.
7. N. M. Batalha, J. Jenkins, G. S. Basri, W. J. Borucki, and D. G. Koch, "Stellar variability and its implications for photometric planet detection with Kepler," in *ESA SP-485: Stellar Structure and Habitable Planet Finding*, B. Battrick ed., Noordwijk, Cordoba, pp. 35–40, 2002.
8. A. C. Robin, and M. Crézé, "Stellar populations in the Milky Way - A synthetic model", *A&A* **157**, pp. 71–90, 1986.
9. M. Haywood, A. C. Robin, and M. Crézé, "The evolution of the Milky Way disc. I. Vertical structure and local constraints," *A&A 320*, pp. 428-439, 1997.
10. M. Haywood, A. C. Robin, and M. Crézé, "The evolution of the Milky Way disc. II. Constraints from star counts at the galactic poles," *A&A 320*, pp. 440–459, 1997.
11. Personal communication, David Koch, 2001.
12. Personal communication, Ron Gilliland, 2003.
13. Personal communication, Neal Nickles, 2003.