The Kepler end-to-end model: creating high-fidelity simulations to test Kepler ground processing

Stephen T. Bryson[a], Jon M. Jenkins[b], Dan J. Peters[c], Peter P. Tenenbaum[d] Todd C. Klaus[d],
Jay P. Gunter[d], Miles T. Cote[a], Douglas A. Caldwell[b],

[a]NASA Ames Research Center, M/S 244-30, Moffett Field, CA USA 94305
[b]SETI Institute/NASA Ames Research Center, M/S 244-30, Moffett Field, CA USA 94305
[c]Ball Aerospace & Technologies Corporation, 1600 Commerce Street, Boulder, CO USA 80301
[d]Orbital Sciences Corporation/NASA Ames Research Center, M/S 244-30, Moffett Field, CA
USA 94305

ABSTRACT

The *Kepler* mission is designed to detect the transit of Earth-like planets around Sun-like stars by observing 100,000 stellar targets. Developing and testing the Kepler ground-segment processing system, in particular the data analysis pipeline, requires high-fidelity simulated data. This simulated data is provided by the *Kepler* End-to-End Model (ETEM). ETEM simulates the astrophysics of planetary transits and other phenomena, properties of the Kepler spacecraft and the format of the downlinked data. Major challenges addressed by ETEM include the rapid production of large amounts of simulated data, extensibility and maintainability.

Keywords: *Kepler*, pixel simulation

1. INTRODUCTION

The *Kepler Mission* continuously observes ∼165,000 target stars in *Kepler's* 115 square degree Field of View (FOV) seeking to discover Earth-like planets transiting solar-like stars by detecting photometric signatures of transits.[1,2] Data is collected and stored for monthly downlink, and the data is processed in the Science Operations Center (SOC).[3,4]

Every target is observed with CCD readout every 6.52 seconds and co-added into 29.4 minute observations, referred to as *Long Cadence* (LC) data. A smaller number of targets, at most 512, is co-added into 58.8 second *Short Cadence* (SC) observations. These data are collected nearly continuously for about 30 days and downlinked via high-bandwidth Ka-band transmissions. Bandwidth constraints of the Ka-band transmission limits the amount data that may be downlinked, making it impossible to downlink all 96 million pixel values that are collected with each LC observation. It is therefore required to identify the pixels that provide the best data for the target stars and select these pixels for downlink. This paper describes the method by which the required pixels for each target are determined.

Kepler operations has three phases: data collection, monthly data downlink and data processing to identify transits and other astrophysical phenomena. These phases involve several organizations that must interface smoothly. ETEM is designed to simulate the collected data, including high-fidelity simulations of astrophysical processes including transit signals and stellar variability, as well as spacecraft noise and systematics. ETEM packages this data in a way that mimics the data as it appears to NASA's deep space network. Therefore ETEM-generated data can be used to test every step of *Kepler* data flow and processing, starting with the arrival of the data on the ground and ending with the identification of planetary transits.

This paper is a description of the further development of ETEM beyond that described in 2004.[12] We briefly review the basic structure of ETEM that is described in that paper and how it has been placed in an extensible software framework.

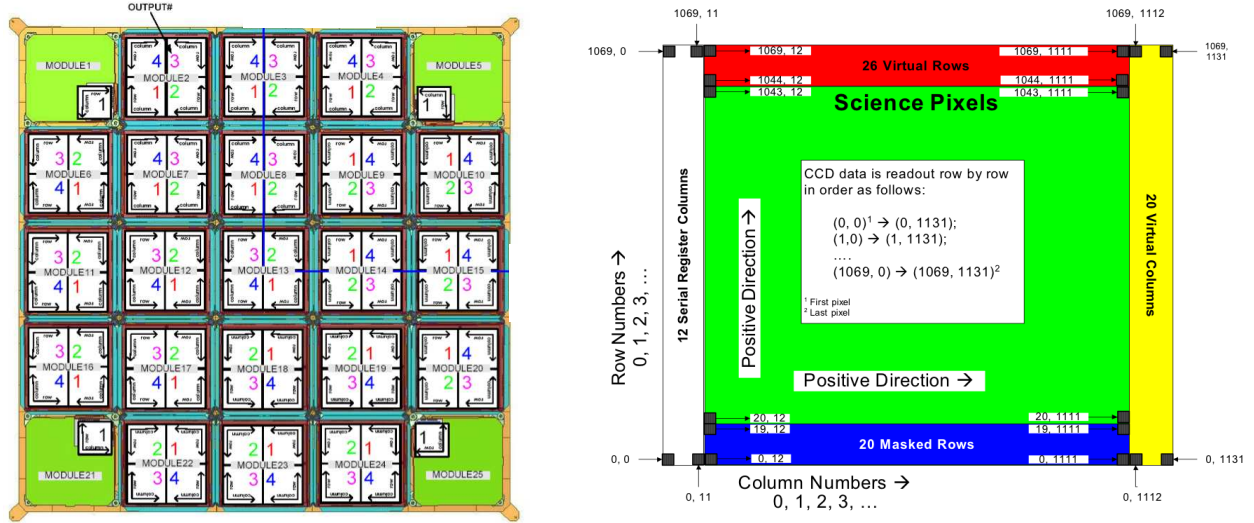Further author information: Send correspondence to S.T.B.: E-mail: Steve.Bryson@nasa.gov

Figure 1. Left: The CCD array on the *Kepler* focal plane, showing the 21 modules, each of which has 2 2200 x 2048 pixel CCDs. Each CCD is read out via 2 output channels. Smaller CCDs used as fine guidance sensors are also shown in each corner, but are not discussed in this paper. Right: the pixel arrangement of each output channel.

1.1 The *Kepler* focal plane

The *Kepler* focal plane science sensors consist of 42 2200 column by 1044 row CCDs mounted on 21 electronic modules (Fig 1) with an image scale of $3.98''$per pixel. The first 20 rows of each CCD are not exposed to the sky in order to provide calibration and diagnostic data as described below. Each CCD is divided into two 1112×1044 *output channels*. Each output channels is supplemented by 26 trailing virtual rows, 12 leading serial register columns and 20 trailing virtual columns, giving each output channel 1136×1070 addressable pixels. The 12 leading serial register columns and 20 trailing virtual columns are used to collect black level data for each row. *Kepler's* lack of a shutter means that pixels are exposed to the sky during readout, which causes image smear along columns. The leading 20 masked and 26 trailing virtual columns measure this smear data. The black level and smear data are called *collateral data* and are used to calibrate the pixel data during ground processing.[7] The pair of CCDs on each module provide a contiguous 2200×2048 pixel image of a portion of the *Kepler* field of view.

Though the Kepler spacecraft is expected to have very high pointing precision, holding star images relatively still on the focal plane, the very wide Kepler field means that differential velocity aberration (DVA) must be modeled. DVA can move a star as much as 0.6 pixels in a quarter.

1.2 *Kepler* pixel and target types

Kepler pixels are collected for several types of targets:

Stellar targets are point-like sources[9] whose pixels are selected to maximize the signal to noise ratio (SNR).[8,14] Stellar targets are specified by a Kepler ID, which is used to look up pertinent data in the Kepler Input Catalog (KIC).[10] Stellar targets may be either LC or SC.

Custom targets are explicitly specified collections of pixels. Custom targets are defined by a reference pixel position and a set of offsets, one for each pixel, from that reference position. Custom targets are used for non-stellar sources and diagnostic collections of pixels, and may be either SC or LC.

Background targets are small (nominally 2×2) sets of pixels that sample the background signal in long cadence. These pixels are selected to support a 2D polynomial representation of the background.[8]

Reference pixel (RP) targets are special stellar targets used for diagnostics whose pixels are downlinked bi-weekly via low-bandwidth X-band communications.[11]

LC, SC, background and RP targets are treated separately by the flight system software, with their own memory allocation. LC, SC and RP share the same mask table, and SC and RP targets also appear on the LC target list. Background targets have a separate mask table.

1.3 The *Kepler* Data Path

Kepler CCD pixel data is converted into a 14-bit digital signal, which is co-added into 23-bit data (embedded in 32-bit words). A nominal long exposure performs 270 co-adds of the 14-bit data. This data can then have several compression options applied. The following describes the nominal path taken by data on the spacecraft:

- Pixel values are requantized into 16 bits via a pre-defined non-linear lookup table that is designed so that the table step sizes are one quarter of the Poisson shot noise that would be associated with the pixel value.

- Every 48 long cadences an uncompressed baseline is stored to support the Huffman encoding in the next section. The difference between this and the previous baseline is taken, and the Huffman encoded differences are also stored to protect against data loss.

- The non-baseline cadence pixel values are subtracted from the latest baseline, and the differences are compressed via Huffman encoding.

- The resulting data are packed into Consultative Committee on Space Data Systems (CCSDS) data source packets and stored in the *Kepler* solid-state recorder.

- For monthly downlink the pixel data are packed first into Virtual Channel Data Units (VCDUs), which are then packaged into Channel Access Data Units (CADUs) which Reed-Solomon encoding, randomization and convolution encoding to increase robustness against data loss during transmission.

- The CADU data is transmitted to the the NASA Deep Space Network (DSN), which unpacks the data into VCDUs which are delivered to the *Kepler* Mission operations center (MOC).

- The MOC further unpacks the data and delivers it to the Data Management Center (DMC) who extracts the actual pixel data values and delivers it as FITS files to the *Kepler* Science Operations Center (SOC) for processing.

- The SOC Data Analysis Pipeline processes the data, calibrating the pixels, performing photometry, searching for planetary transits and other astrophysical phenomena and monitoring spacecraft health.

ETEM is tasked with generating high-fidelity synthetic data that exercises every step in the above chain with realistic simulations of *Kepler* pixel data containing expected astrophysical effects including planetary transits and spacecraft noise and systematics. In the above chain of data, the simulated pixel value up to requantization into 16 bits is implemented in MATLAB, and the Huffman encoding and packaging of the data is implemented in java. ETEM is designed to be extensible in several ways, allowing increased knowledge of the spacecraft and astrophysics to be inserted as needed.

2. EFFICIENTLY SIMULATING ASTROPHYSICAL PHENOMENA

An ETEM simulation is performed per-output channel, and takes as input a variety of data:

The *Kepler* Input Catalog[10] (KIC) providing information about stars in the *Kepler* field.

The Pixel Response Function[14] (PRF), an observation-based super-resolution representation of how starlight falls on pixels. The PRF includes the optical point spread function convolved with intra-pixel variability and high-frequency pointing jitter.

Target definitions[?,?] that define the target stars and which pixels are to be observed in LC, SC and RP.

Model Solar-like variability

Pointing jitter modeltadSPIE, an estimate of the low-frequency spacecraft pointing jitter.

Focal plane geometry (FPG) and pointing model,[15] which includes measurements of the locations of the CCDs in the *Kepler* focal plane, models of the Kepler optics and of DVA. These models are used to determine the pixel location of the central ray of each stellar target.

Saturation model,[15] which includes information about the well depth of each output channel.

Other data about CCDs and system electronics,[15] such as flat fields, CCD charge diffusion and charge transfer efficiency (CTE), electronic dark levels, observed instrumental noise.

2.1 Simulating stars, their positions and motions

A primary concern of ETEM is the ability to produce 90 days of simulated data for all 84 channels in a reasonable amount of time. This is facilitated by parallelizing the simulation, computing each output channel's data on a separate system in the SOC cluster.[3] But a single channel contains tens of thousands of stars in the KIC and $\sim$ 2000 observational targets that require dynamic modulations.

The motion of all simulated stars due to DVA and pointing jitter must be modeled by ETEM. Rather than rendering all stars on an output channel from scratch, which would be quite slow, ETEM develops a linear polynomial model of the response to each pixel's starlight to motion of that star on a sub-pixel (nominally 0.1 pixel) grid. This approach takes advantage of the highly optimized linear algebra algorithms used by MATLAB.

First, each star in the KIC that falls on the output channel being simulated is projected onto the sub-pixel grid. Then a polynomial representation $PRF(\Delta x, \Delta y)$ of the PRF for this channel as a function of offset $(\Delta x, \Delta y)$ is created on the sub-pixel grid that covers the extent of DVA and jitter motion. For each sub-pixel position, the flux (as determined from the KIC) of all stars projected on that position is summed and then convolved with each coefficient of $PRF(\Delta x, \Delta y)$. The result is a representation of the flux in pixel (r, s) of the form

$$p_{r,s}(\Delta x, \Delta y) = \sum_{i,j=0}^{O} c_{i,j} \Delta x^i \Delta y^j. \tag{1}$$

Here the coefficients $c_{i,j}$ is the convolution of the flux falling on pixel (r, s) with the corresponding coefficient of $PRF(\Delta x, \Delta y)$, and O is the number of coefficients as determined by the order of the polynomial. The result is a set of polynomial coefficients for all pixels on the output channel pixel array, which can be quickly evaluated for any small offset $(\Delta x, \Delta y)$. For details see the 2004 ETEM paper.[12]

Signal modulations such as stellar variability and transit signals are modeled only on target stars.

LC compared with SC, RP

2.2 Transit simulation

Describe the implementation of Mandel and Agol's algorithm for both binary stars and for planets

2.3 Cosmic rays and other astrophysics

3. SIMULATING SPACECRAFT AND INSTRUMENTAL EFFECTS

3.1 Focal plane models

flats (small and large scale) saturation

3.2 Observed systematics

FGS clocking crosstalk example

4. A FRAMEWORK TO SUPPORT EXTENSIBILITY

5. TESTING EXPERIENCE

6. CONCLUSIONS

ACKNOWLEDGMENTS

REFERENCES

[1] Borucki, W., et al., "Kepler planet-detection mission: introduction and first results, *Science*, **327**, pp. 977–980, 2010.

[2] Koch, D. G. *et. al.*, "Kepler mission design, realized photometric performance, and early science," *ApJ Lett.*,**713(2)**, pp. L79–L86, 2010.

[3] Middour, C. K. *et. al.*, "Kepler Science Operations Center Architecture," these proceedings, 2010.

[4] Haas, M. *et. al.*, "Kepler Science Operations," ApJL 713 (2010) L115-L119.

[5] Van Cleve, J., & D. A. Caldwell, 2009, Kepler Instrument Handbook, KSCI 19033-001, (Moffett Field, CA: NASA Ames Research Center)

[6] Caldwell, D. A. *et al.*, "Instrument Performance in Kepler's First Months," ApJL 713, L92-L96.

[7] Quintana, E. V. *et. al.*, "Calibrations," these proceedings (2010).

[8] Bryson, S. T. *et al.*, "Selecting Pixels for Kepler Downlink," these proceedings (2010).

[9] Batalha, N. M. *et. al.*, "Selection, Prioritization, and Characteristics of Kepler Target Stars," *The Astrophysical Journal Letters*, **713**, pp. L109–L114, 2010.

[10] Latham, D. W. *et. al.*, "The Kepler input catalog," Proc. AAS 207, 1340 (2005).

[11] Jenkins, J. M. *et al.* "Semi-weekly monitoring of the performance and attitude of Kepler using a sparse set of targets," these proceedings, 2010.

[12] Jenkins, J. M. *et al.* "An Efficient End-To-End Model for the Kepler Photometer," SPIE 5497 2004.

[13] Mandel, K. and Agol, E., "Analytic Lightcurves for Planetary Transit Searches," *The Astrophysical Journal*, **580**, pp. L171-L175, 2002.

[14] Bryson, S. T. *et al.* "The Kepler Pixel Response Function," *The Astrophysical Journal Letters*, **713**, pp. L97–L102, 2010.

[15] Allen, C. A. *et. al.*, "Focal Plane Characterization," these proceedings (2010).

[16] Twicken, J. D. *et al.*, "Photometric Analysis in the Kepler Science Operations Center Pipeline ," these proceedings (2010).