

# The Little Photometer That Could: Technical Challenges and Science Results from the *Kepler* Mission

Jon M. Jenkins<sup>°a</sup>, Jeb Dunnuck<sup>b</sup>,

<sup>a</sup> SETI Institute/NASA Ames Research Center, M/S 244-30, Moffett Field, CA USA 94305

<sup>b</sup> Ball Aerospace & Technologies Corporation, P.O. Box 1062, Boulder, CO, USA 80306

## ABSTRACT

The *Kepler* spacecraft launched on March 7, 2009, initiating NASA's first search for Earth-size planets orbiting Sun-like stars. Since launch, *Kepler* has announced the discovery of 17 exoplanets, including a system of six transiting a Sun-like star, Kepler-11, and the first confirmed rocky planet, Kepler-10b, with a radius of 1.4 that of Earth. *Kepler* is proving to be a cornucopia of discoveries: it has identified over 1200 candidate planets based on the first 120 days of observations, including 54 that are in or near the habitable zone of their stars, and 68 that are 1.2 Earth radii or smaller. An astounding 408 of these planetary candidates are found in 170 multiple systems, demonstrating the compactness and flatness of planetary systems composed of small planets. Never before has there been a photometer capable of reaching a precision near 20 ppm in 6.5 hours and capable of conducting nearly continuous and uninterrupted observations for months to years. In addition to exoplanets, *Kepler* is providing a wealth of astrophysics, and is revolutionizing the field of asteroseismology. Designing and building the *Kepler* photometer and the software systems that process and analyze the resulting data to make the discoveries presented a daunting set of challenges, including how to manage the large data volume. The challenges continue into flight operations, as the photometer is sensitive to its thermal environment, complicating the task of detecting 84 ppm drops in brightness corresponding to Earth-size planets transiting Sun-like stars.

**Keywords:** *Kepler* Mission, exoplanet, transit, photometry, data compression, systematic error correction

©2011 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

## 1. INTRODUCTION

The *Kepler* spacecraft was launched in March 2009 as NASA's first mission capable of finding Earth-size planets orbiting Sun-like stars. This mission has been a spectacular success in terms of revolutionizing both the fields of extrasolar planets and of asteroseismology, the study of stellar oscillations. Over 1200 planetary candidates were detected in the first 120 days of observations,<sup>1</sup> effectively tripling the number of known extrasolar planets as of February 2011. Most of these candidates are as small or smaller than Neptune. *Kepler* finds multiple planet candidates transiting a star ~17% of the time. With the capability to conduct observations for at least 10 years and a photometric precision approaching 20 ppm on 6.5-hr timescales, *Kepler* is poised to reach its primary goal of determining the frequency and distribution of rocky planets in the habitable zone (HZ) of solar-like stars, that range of distances for which liquid water could pool on the surface of a terrestrial planet. However, the success of the *Kepler* Mission wasn't always a foregone conclusion.

This paper describes some of the major challenges facing *Kepler* both before and after launch. These challenges have been mitigated by a combination of good design, great hardware, flexible software and the talented people involved in the mission. It required collaboration between the engineers and scientists who designed and built the photometer and spacecraft and those who process and analyze the exquisite photometric data. We focus on some of the key challenges that were met handily by the design, including the performance of the onboard compression scheme and the approach to target management, and some issues that have arisen after launch when we had the first opportunity to see how a photometer such as *Kepler* would perform in the space environment. It's true that these one-of-a-kind machines never cease to amaze or surprise us when they are finally in orbit and operating, in part, because we cannot truly replicate the totality of the space environment in the laboratory. Also, we cannot ever anticipate the myriad complexity of nature when we observe

---

<sup>°</sup>Send correspondence to J.M.J.: E-mail: Jon.Jenkins@nasa.gov

it with a new and much more powerful instrument than has been previously available. We give examples of how we are dealing with some of the post-launch complexities in ground-processing software, including the thermal sensitivity of the photometer. We provide an update on the key scientific achievements obtained with *Kepler* to date. Finally, we conclude with some thoughts on what the future holds for *Kepler*, and how our experience with this marvelous machine can provide lessons for future space missions, such as ESA's envisioned PLATO space transit survey.

## 2. KEY CHALLENGES

In 1994, the first proposal for *Kepler* to NASA's Discovery Program was turned down as too expensive and too risky from a technical perspective. In fact, *Kepler* was rejected by the next two Discovery panels in 1996 and 1998, but was ultimately selected for launch in the fourth Discovery Program competition in 2001 after several years of dedicated and intense efforts to demonstrate the technical and budgetary feasibility of the mission in the face of many significant technical challenges.

*Kepler*'s goal is simple: to observe a large sample of solar-like stars (those burning hydrogen in their cores) to detect periodic dimming in brightness caused when a small planet transits, i.e., crosses the face of its star as it swings around in its orbit, as viewed by the spacecraft. There are several complications with this approach:

First, the dimming caused by a small planet transiting its star is very slight. Planets are rather small compared to the size of stars, even main-sequence stars, e.g., our Sun, which are called dwarfs in comparison to their older, evolved brothers, the giant stars, which burn helium in their cores. The amplitude of the dimming during transit is the ratio of the area blocked by the planet to the total area of the star. Jupiter is about 10% the radius of the Sun, so its transits block 1% of the light as viewed by an extrasolar observer. By comparison, Earth is about 10% the radius of Jupiter and consequently causes a 0.01% or 1 part in 10,000 drop in brightness when it transits the Sun. Since the opportunity to detect a signal arises only when the signal is strong compared to the observational noise, robustly detecting an Earth-size planet transiting a Sun-like star requires a signal-to-noise ratio (SNR) of at least  $4\sigma$  over the duration of a typical transit,  $\sim 6.5$  hours in order to build up a total  $\text{SNR} > 8\sigma$  from four or more transit events. The threshold of  $7.1\sigma$  is required in light of the large number ( $\sim 10^{12}$ ) of statistical tests performed to identify the transit sequences, since we don't know the period or the time of first transit.<sup>2,3</sup> In the 1990s, the precision achieved by state of the art differential photometry from the ground with Charge Coupled Devices (CCDs) was  $\sim 0.1\%$ , and could not be achieved easily on a routine basis even with 4-m class telescopes due to atmospheric effects such as extinction, and scintillation,<sup>4</sup> which continue to limit the precision of ground-based photometry.

The *Kepler* concept called for balancing the three major sources of noise to achieve a single transit SNR of  $4\sigma$  for an 84-ppm, 6.5-hr transit of a G2V star by an Earth-size planet in a one-year orbit. This point design established the three major noise sources and their required values as: 1) shot noise from Poisson photon-counting statistics ( $\leq 14$  ppm), instrument noise ( $\leq 10$  ppm), and stellar variability ( $\leq 10$  ppm).<sup>5</sup> A major question raised by the first three Discovery review panels was whether CCDs could achieve an instrumental precision of 10 ppm or less, and further, whether a photometer composed of CCDs operating in a space-like environment could achieve this goal. After all, even though a telescope in deep space avoids the problems of peering up through the atmospheric soup surrounding the Earth, there are instrumental effects and noise even in the space environment. A series of laboratory measurements established that front-side illuminated CCDs can achieve the precision necessary for *Kepler*.<sup>6</sup> The arrival of readily available backside-thinned, illuminated CCDs brought the quantum efficiency of the proposed detectors to  $> 90\%$ , enhancing the prospects for *Kepler*, as this improvement significantly reduced the shot noise in measuring the brightness of star images. Laboratory studies with backside-illuminated CCDs demonstrated that they, too, could be operated in such a way as to achieve several ppm of instrumental noise on transit timescales,<sup>7</sup> and that they experienced less sub-pixel variations in sensitivity. Finally, the *Kepler* Technology Demonstration (KTD), funded and conducted after the 1998 Discovery proposal cycle, demonstrated that the *Kepler* design could reach the requisite instrumental precision in the face of a variety of disturbances of importance to a space-borne photometry mission.<sup>8,9</sup> These included residual pointing errors due to the rumble of reaction wheels used to control pointing and the band-limited performance of the attitude control system. Cosmic rays, bright, saturated stars in the field of view, and focus changes due to changing thermal conditions were also simulated. The KTD also provided direct simulations of Earth-size transit signals through the application of a small current through thin, ribbon wires placed across rectangular pinholes in the opaque plate that were imaged onto the test CCD\*, an EEV 42-80<sup>†</sup>. A few milliamperes of

\*An ingenious concept due to David Koch, Deputy Science Principal Investigator.

<sup>†</sup>The E2V CCDs built and flown on *Kepler* were derived from the EEV 42-80 CCD.

current heated the wires, causing them to expand by about one atom's width, thereby blocking  $\sim 0.01\%$  of the rectangular pinhole. The results allowed us to demonstrate that we could detect Earth-size transit signals against the shot noise and instrumental effects we knew would be at play in the space environment.

Second, the timescale and low duty cycle of transit signatures present serious challenges to Earth-bound transit photometry campaigns: a central transit of an Earth-size planet of a Sun-size star in an Earth-size orbit (1 AU) lasts for  $\sim 13$  hours, and repeats only once per year. It is difficult to imagine achieving measurements of a quality and continuity from the ground required to reliably observe, let alone, detect transits of an Earth-size planet, even if the requisite photometric precision could be routinely obtained. It is true that prior to *Kepler's* launch,  $\sim 70$  planets were found that transited their parent star. Almost all of these are at least as large or larger than Jupiter, providing transit depths of  $\sim 1\%$ , and in very short period orbits of only a few to several days. This set of circumstances allows for many transits during observing seasons lasting a few months, and transit durations lasting from one to a few hours, permitting detection from the ground. The fact that the stars can only be observed in visible wavelengths at night, the unpredictable nature of weather (which often interferes with astronomical observations), and the challenges in scheduling coordinated observations at multiple telescopes distributed around the globe all contribute to the difficulty of detecting habitable Earth-size planets from the ground<sup>‡</sup>.

Third, the geometrical probability that the orbit of an extrasolar planet will be aligned so that we can observe its transits is small for an Earth-like orbit: the probability is the ratio of the size of the star to the size of the orbit. Thus, observers in our solar neighborhood have a 0.5% chance of seeing the Earth transit the Sun. This implies that for a robust null result, *Kepler* needs to observe thousands of stars in order to place strong upper limits on the frequency of such planets in the case that none are detected.

Fourth, we don't know when a star is going to dim due to a transit, and transits of terrestrial bodies transpire on timescales of 1–16 hours. The duty cycle of a transiting Earth-analog is only 0.15%. *Kepler* counts on building the SNR of a transit signature by combining together separate events and will be observing a handful of transits for Earth-analogs.

The solution to these challenges is to launch a photometer with a sufficiently sized aperture into a heliocentric, Earth-trailing orbit with a field of view large enough to observe a sample of  $\geq 100,000$  solar-like stars with instrumental “*Kepler*” magnitudes,  $K_p$  from 9 to 15 as continuously and contiguously as possible.<sup>5,10</sup> The design and build-up of the focal plane array, the largest ever built and launched by NASA, and the photometer itself were extremely challenging endeavors in the cost- and schedule-constrained aegis of NASA's Discovery Program.<sup>11,12</sup> *Kepler's* Earth-trailing orbit provides a quiescent and stable environment in which to operate. The spacecraft's thermal environment is largely controlled by the measured and stately progress of the apparent position of the Sun moving along the ecliptic from *Kepler's* point of view. The greatest force on the spacecraft is the torque due to the force of the photon pressure on the spacecraft and instrument. The Schmidt design of the photometer, a telescope with a 0.95-m aperture and a 1.4-m primary mirror, coupled with a focal plane that consists of 42 CCDs containing 94,617,600 pixels spanning 115 square degrees of sky, allow *Kepler* to simultaneously and nearly continuously observe  $\sim 165,000$  stars. The target stars were selected using the Kepler Input Catalog (KIC) to identify the stars for which we can most readily detect and confirm transits of small rocky planets.<sup>13,14</sup> The Attitude Determination and Control System (ADCS) provides extremely precise pointing with a required and delivered performance of *rms* pointing errors at the level of 0.009 arcsec,  $3\text{-}\sigma$  on time scales upwards of 15 minutes. As the pixels span 3.98 arcsec on each side, this translates to pointing precision of 0.75 millipixels,  $1\text{-}\sigma$ . This is  $\sim 1000\times$  better than that of the Hubble Space Telescope on a pixel scale. The compression scheme and the onboard storage allow *Kepler* to observe continuously for month-long intervals with brief interruptions to allow for downlinking the data to the ground.

The following sections describe some of the key challenges related to collecting and processing the science data and our efforts to mitigate them. Section 3 addresses the key challenge of monitoring the target stars as continuously as possible by minimizing both the number of planned interruptions to the observations as well as the time it takes to downlink the data from the spacecraft. Section 4 describes key challenges posed by nature in the form of stellar variability and by the subtle, but very important, thermal sensitivity of the instrument.

---

<sup>‡</sup>It appears to be impractical to detect Earth-size planets in the HZ of Sun-like stars by any means from the ground, given the small mass and size of such objects, which is why there are space missions proposed to detect them by transit, by ultra-high contrast imaging and by gravitational microlensing.

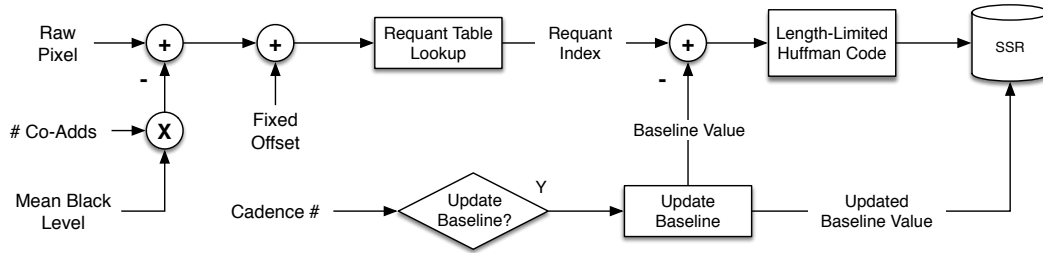


Figure 1. The flow of the compression algorithm implemented in flight software. The co-added pixel values are first shifted to a common zero point through the use of a fixed offset and the mean black level (bias). The shifted pixels are then mapped to the nearest neighbor in the requantization table that has larger steps between large pixel values and smaller steps between small pixel values so that the quantization noise is a fixed fraction of the intrinsic measurement uncertainty. A daily baseline value is subtracted and the residuals are entropically encoded using a length-limited Huffman code in order to squeeze out temporal redundancies in each pixel time series. The compressed bit stream is written to the Solid State Recorder for later downlink to the Deep Space Network.

### 3. MANAGING *KEPLER*'S DATA VOLUME

One of the key challenges for *Kepler* is the management of the data volume. A full frame image (FFI) for *Kepler* has 101,744,160 pixels, including 7,126,560 collateral pixels used for calibration read out from 84 CCD module outputs (2 per CCD). The individual 6.54-sec frames are co-added for 29.4 minutes (a Long Cadence interval). Consequently, the full, uncompressed image data acquired every day are >300 GB. The full frame images contain significantly more data than we can afford to store onboard the Solid State Recorder (SSR), which must be able to hold at least two months of data for robustness against a missed monthly downlink opportunity. The strategy for reducing the data volume consists of first retaining only the pixels of interest (POI) and then applying a lossless compression scheme to squeeze out the temporal redundancies in the data stream. The responsibility for managing the data volume is distributed between the Science Operations Center (SOC) at NASA Ames Research Center, and *Kepler*'s flight software. The SOC provides the compression tables and the target definitions, and the flight software uses this information to select the pixel data that are compressed and stored on the SSR. The Data Management Center at STScI decompresses the data on the ground prior to sending them to the SOC. The success of this approach depends on a number of factors.

#### 3.1 Discarding Pixels of Little or No Interest

The first step in reducing the data volume is accomplished through the use of target definitions, which allow us to specify up to 170,000 Long Cadence (LC) stellar targets for 29.4-minute integrations which can occupy up to 5,440,000 pixels. Each definition contains a reference row and column along as well as an index into an aperture mask table onboard *Kepler* containing 1024 different masks that vary in shape and size.<sup>15</sup> 512 Short Cadence (SC) targets with up to 43,520 pixels can also be specified for 58.7-sec integrations<sup>§</sup>. The KIC, together with models for the pointing behavior of the spacecraft and the point spread function (PSF),<sup>16</sup> allows us to identify the pixels required for each stellar target based on an SNR analysis of the average scene surrounding each target star.<sup>15</sup> More pixels are assigned to bright stars than to dim stars: LC targets are allocated an average of 32 pixels per star, while SC targets are allocated an average of 85 pixels per star. The aperture masks assigned to the targets don't contain much margin for error: a one-pixel halo is added to the pixels identified as essential to calibrating and making the photometric measurements. The masks are efficient: only 4% of the pixels addressed by the masks are not pixels of interest.<sup>15</sup> The POI include sufficient pixels to monitor the background flux levels (up to 4500 per CCD output) and to allow us to correct for on-chip artifacts such as variations in the bias voltage added to the CCD output prior to digitization, and to monitor and remove the effects of the shutterless readout smear which occurs because *Kepler* operates without a mechanical shutter.<sup>17,18</sup> No background pixels are collected at SC, and the smear and black (bias) collateral data are only collected at the SC intervals for the rows and columns containing SC target pixels.

The success of this approach to reduce the amount of data stored and transmitted by *Kepler* relies on several factors. We need to understand in great detail the scene around each potential target star so that we can render accurate synthetic images for the SNR calculation supporting the identification of the required pixels. This is supplied by the KIC, which contains

<sup>§</sup>Each Long Cadence interval consists of 30 Short Cadence intervals, and each Short Cadence consists of 9 co-added 6.54-sec integrations.

$4.5 \times 10^6$  stars and their characteristics<sup>¶</sup>. Generating the photometric scene for each target star requires knowledge of the PSF, and knowledge of exactly where the star is going to be throughout the three months of upcoming observations. The Pixel Response Function (PRF), essentially a 2-D representation of the interaction of the PSF with the sub-pixel sensitivity variations in the CCDs, allows the stellar images to be rendered accurately. How the celestial coordinates of the stars map to silicon coordinates is defined by the Focal Plane Geometry (FPG) model. These models are coupled to a heuristic, predictive model of the pointing behavior of the spacecraft in order to identify the POI. Obtaining the PRF and FPG models was a major feat and resulted from a set of unique data collected and analyzed during the commissioning phase.<sup>16, 19, 20</sup> A total of 121 15-minute LC frames were collected over a grid of 2-D offsets over  $\pm 0.5$  pixels. These observations were made for 52,496 stars and used over-sized aperture masks with two halos to allow for uncertainties in the pre-launch FPG models. Approximately 200 of the brightest, unsaturated stars on each CCD readout channel were used to simultaneously solve for the super-resolution PRFs, which were allowed to vary in shape over each CCD readout channel, and for the FPG coefficients. The pointing model which plays an essential supporting role in these calculations assumes that once the ADCS locks onto the guide star, it acts to minimize the movement of the guide stars from their starting positions. This simple, heuristic pointing model has proven to be quite successful despite the fact that the detailed motion of the stellar images cannot be modeled by a static, rigid-body coordinate transform, due to small, thermally induced, non-uniform focus changes. The Target and Aperture Definition (TAD) pipeline component uses these models to analyze the average SNR of each target's flux measurement over the upcoming three months of observation. Note that differential velocity aberration causes the stars at the edges of the large  $\sim 13.25^\circ \times 13.25^\circ$  field of view to move by up to  $\sim 0.6$  pixels over 90 days so the aperture masks must be large enough to capture this motion.

Of equal importance to this task is the pipeline component, Photometer Data Quality (PDQ), which reconstructs the attitude of the spacecraft after each return to science data collection to confirm that each and every target star is within  $\pm 0.1$  pixels or 0.398 arcsec of its expected location.<sup>21</sup> If we can't achieve this accuracy then the target definitions will become invalid during the observations and we lose essential data. This is a daunting task: the focal plane is approximately 12,000 pixels across. The required precision is therefore 1 part in 120,000. This is comparable to the photometric precision required to detect Earth-size transits. If any target star falls outside this very small bull's-eye, then PDQ furnishes a pointing "tweak" in the form of a delta quaternion that represents the error between the desired pointing and the actual pointing reconstructed from the data. Upon request by the *Kepler* Science Office, this delta quaternion can then be uplinked and injected into the ADCS as an artificial pointing error that will be nulled out and therefore, nudge the spacecraft to the correct attitude. PDQ itself has far fewer pixels to work with than is available for normal science, as its data is provided through the low gain antennas at X-band. At the start of the mission, PDQ relied on no more than 96,000 pixels with about 300 star images from the focal plane array, and now only has access to  $\sim 160$  stars. The number of stellar targets that PDQ receives for its attitude reconstruction will continue to drop as it drifts away from Earth and the strength of the radio link to the spacecraft diminishes. So far, PDQ has proven itself up to the task.

The use of the target definitions reduces the fraction of pixels that are stored on the SSR to  $\sim 5\%$  of that in a FFI. The raw data are represented as 23 bit words and the compressed data must occupy no more than 5.75 bits per pixel on average in order for the SSR to achieve adequate storage capacity. The uncompressed POIs are still a factor of 4 too large to store on the SSR. We can take advantage of two aspects of the digitized data to achieve the necessary level of compression: 1) the digitized pixel data for bright pixels are highly over-represented, and 2) the temporal redundancy in the observations. The former is handled by the requantization scheme described in Section 3.2, and the latter is addressed by a length-limited Huffman code discussed in Section 3.3.

### 3.2 Requantization

First, the 14-bit Analog-Digital-Converters (ADC) quantize the pixel measurements uniformly to  $\approx 100 \text{ e}^-/\text{ADU}$ , regardless of the intrinsic measurement noise of each observation. As a consequence, the bright pixels are over-represented due to shot noise, which varies as the square root of the number of photons as an *rms* value. For example, the well depth of the CCDs is approximately  $1 \times 10^6 \text{ e}^-$ . Thus, after 270 exposures (an LC interval), the brightest pixels have  $270 \times 10^6 \text{ e}^-$  with a corresponding shot noise of  $1,6432 \text{ e}^-$ , and the quantization is at the  $100 \text{ e}^-$  level,<sup>22</sup> so these pixels have a resolution that is 164 times smaller than the intrinsic measurement uncertainty. To reduce the degree to which bright pixels are over-represented, *Kepler* applies a software "compander" to redistribute the bits associated with the pixel measurements

---

<sup>¶</sup>Generating the KIC was a significant challenge in itself and required several observing seasons and a significant amount of effort to reduce the data and characterize the stars.

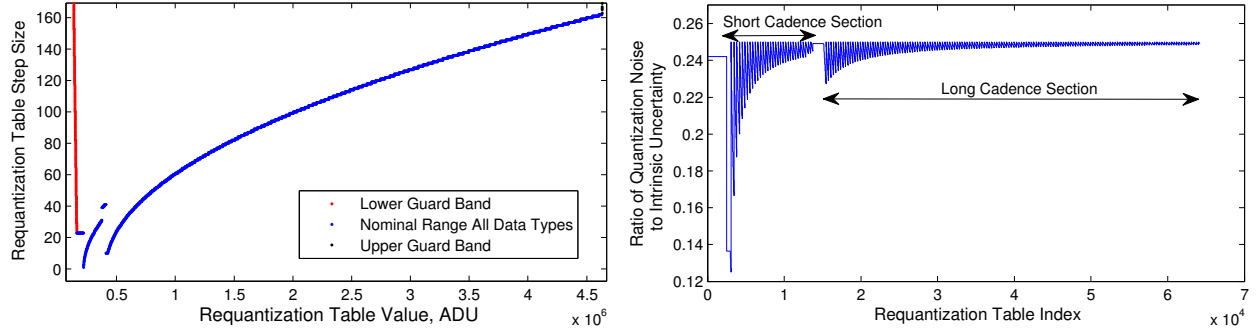


Figure 2. Left: The requantization table step sizes as a function of table index. The two main photometric data sections are shaped approximately like the square root function, with low middle and high guard band sections above, below and in between the short cadence and long cadence sections of the table. Right: A verification figure showing that all data types are quantized to no more than  $1/4$  of their intrinsic uncertainties.

so as to make the ratio of the quantization noise to the total intrinsic noise a fixed value chosen to be  $r = 1/4$ , called the requantization ratio. This is accomplished with a “requantization” table that is a list of allowed pixel values whose spacing increases as a function of the measured pixel value. The requantization algorithm replaces each pixel measurement with the index of the nearest pixel value in the requantization table (containing exactly  $2^{16}$  entries) by the flight software, i.e., it is a simple look-up table with nearest neighbor interpolation.

What effect does requantization have on the total noise? We have

$$\sigma_{\text{total}}^2 = \sigma_{\text{intrinsic}}^2 + \Delta_Q^2/12 = \sigma_{\text{intrinsic}}^2 + r^2 \sigma_{\text{intrinsic}}^2 = (1 + r^2) \sigma_{\text{intrinsic}}^2 \quad (1)$$

where  $\sigma_{\text{total}}^2$  is the combination of the measurement uncertainty,  $\sigma_{\text{intrinsic}}^2$ , and the quantization noise,  $\Delta_Q^2/12$ , due to a quantization step size of  $\Delta_Q$ .<sup>23</sup> The requantization increases the *rms* noise on the observations by  $\sigma_{\text{total}}/\sigma_{\text{intrinsic}} - 1 = \sqrt{1 + r^2} - 1 = 0.03$  for the choice  $r = 1/4$  a rather modest increase of 3%. This is acceptable as this approach treats the quantization noise as a term in the noise budget and the requantization scheme allows us to trade a slight increase in the noise level of the photometric measurements for a large reduction in the data volume taken up by the measurements and the time required to downlink the data.

How do we build the requantization table? The concept is simple: start with the lowest photoelectron count expected in terms of digital counts,  $p_{\text{min}}$ , and set the first value in the requantization table,  $p_1$  to this value. Set the counter  $i = 1$ . The intrinsic uncertainty in a pixel value,  $p_i$ , in ADU is given by

$$\sigma_{p_i}^2 = (p_i - n_{\text{co-adds}} \cdot \mu_{\text{black}}) / G_j + n_{\text{co-adds}} \cdot \sigma_{\text{read}}^2, \quad (2)$$

where  $n_{\text{co-adds}}$  is the number of individual exposures that were summed to obtain the pixel value, including any spatial co-adds,  $\mu_{\text{black}}$  is the mean black or bias level in ADU,  $G_j$  is the gain (ADU/e<sup>-</sup>) for the CCD channel  $j$ , and  $\sigma_{\text{read}}$  is the *rms* read noise in ADU (typically 1 ADU). Compute the step size for element,  $i$ , as

$$\Delta_Q(p_i) = \max \left\{ 1, \left\lfloor \sqrt{12} r \sigma_{p_i} \right\rfloor \right\}, \quad (3)$$

where  $\lfloor \cdot \rfloor$  is the floor operator. Then  $p_{i+1} = p_i + \Delta_Q(p_i)$ .

In practice, generating the requantization table for the flight segment is complicated by the fact that the bias voltages added to the CCD readouts vary significantly across the focal plane, and they vary spatially even within each CCD readout channel.<sup>22</sup> The SOC furnishes a mean black table to the flight segment for this use in addition to the requantization table. The electronic gains also vary significantly over the 84 CCD readout channels.<sup>22</sup> Moreover, the collateral data used to calibrate the photometric data for on-chip artifacts, such as the bias voltage (or black level) and the shutterless readout smear, are co-added spatially before being compressed. Also, there are two cadences for photometric data, Short Cadence, consisting of 9 summed readouts, and Long Cadence, consisting of 270 summed CCD readouts. The readout noise associated with SC and LC are quite different for pixels having the same value. Because of this, it is advantageous to

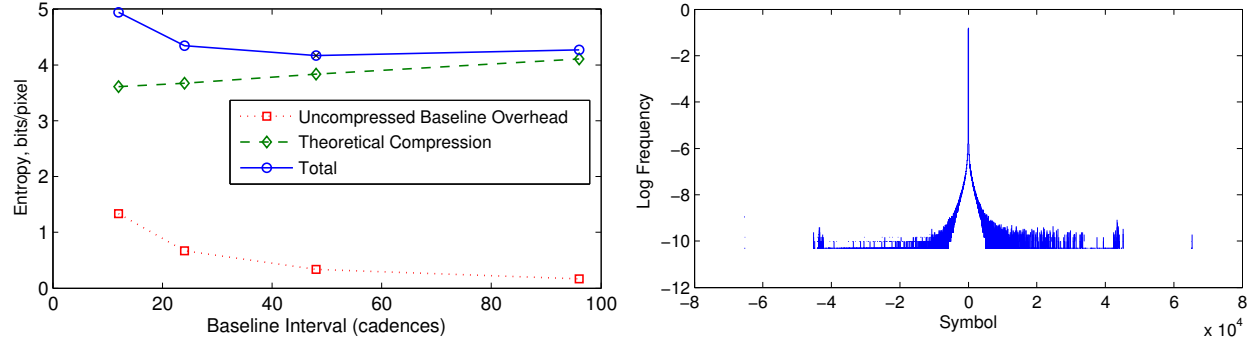


Figure 3. Left: The compression rate as a function of the baseline interval for the predictor for four different baseline intervals. The solid blue curve with circles is the total compression rate, while the dashed green curve with diamonds is the entropy for the compressed data and the dotted red curve with squares is the overhead cost associated with storing the baselines, which are not compressed. There is a shallow minimum at 48 cadences ( $\sim 1$  day), which is the value selected for the baseline interval. Right: Histogram of the symbol frequencies corresponding to a baseline interval of 48 Long Cadences. The symbols can span the range from  $-2^{16} + 1$  to  $2^{16} - 1$ . The compactness of the residuals dictates the compression that can be achieved. Over 99% of the pixel residuals are expected to be within 13 requantization levels of the daily baseline value.

design separate requantization table sections for the SC and LC data to help maximize the compressibility of the resulting data. The SOC generates a table combining the LC and SC sections, and uses a LC fixed offset of 419,400 ADU, and a SC fixed offset of 219,400 ADU that allow the flight software to tailor the requantization to the cadence type for each pixel measurement. Generating the requantization table requires a bit of bookkeeping.

Equations 2 and 3 are exercised for each table section to calculate the step size required by each SC or LC data type across the focal plane. The step size taken is the smallest one required, after which the process repeats until  $p_i$  exceeds the expected range of the photometric measurements. This design allows for guard bands above, below, and between the two main requantization tables to ensure that drifts in the DC bias voltages don't result in variations in pixel values that fail to be captured by the requantized data because the input values fall outside the expected input range. Requantization reduces the size of the data from 23 bits to 16 bits per pixel per cadence, a substantial, but insufficient reduction in the data volume.

### 3.3 Length-Limited Huffman Coding

The second part of the compression scheme takes advantage of the slowly varying nature of the scene imaged by *Kepler* to reduce the variance in each POI's time series by removing a predicted value for each POI. This prediction is furnished by a requantized measurement for each pixel made during a baseline cadence interval, and this baseline measurement is updated every 48 cadences (which is a software parameter) starting with the first cadence collected after return to science mode. The residuals from the baseline value are entropically encoded using a length-limited Huffman code<sup>24–26</sup> that limits the code words to no more than 24 bits. Huffman codes compress data streams by representing common symbols with short binary code words and uncommon symbols with long code words<sup>†</sup>. Length-limited Huffman codes are somewhat less efficient than non-length-limited codes, but the performance in the context of *Kepler* is more than adequate. The uncompressed baseline pixel values are stored onboard the SSR as well as the differences between the previous baseline value and the new baseline value for each pixel, which is compressed. This allows us to reconstruct the requantized pixel values from the bit stream both forward from the previous baseline as well as backwards from the following baseline in case a baseline measurement is lost or corrupted.

Generating a Huffman coding table requires that the frequencies of the symbols to be coded be quantified. In *Kepler*'s case, we needed to generate and fly a compression table prior to flight for test purposes. To do so, we generated high fidelity, synthetic science data using the *Kepler* End-To-End Model (ETEM)<sup>27,28</sup> and ran it through the Huffman Generator (HGN) pipeline where the symbol histograms are generated for each of the 84 CCD readout channels\*\* for the list of baseline

<sup>†</sup>Huffman codes are ubiquitous and are used in various digital coding standards including mp3 and jpeg applications as well as in fax machines.

\*\*Each CCD has two readout channels, or module outputs.

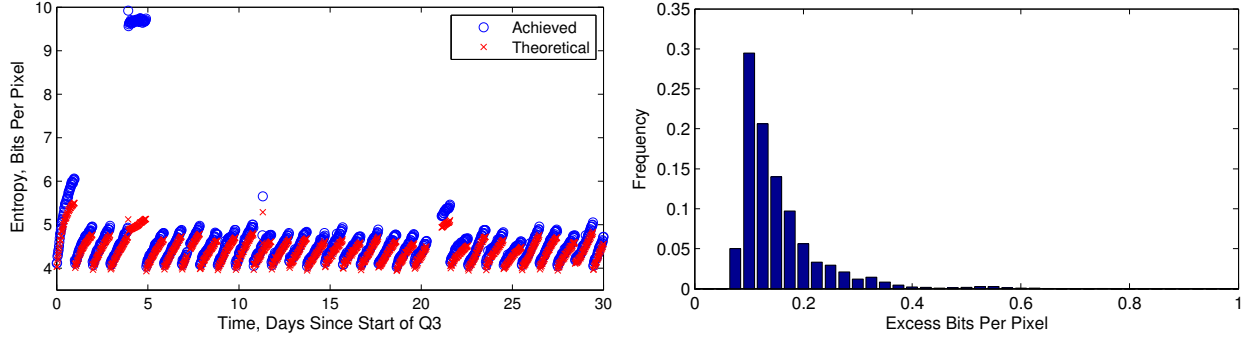


Figure 4. Left: Achieved and theoretical compression rates for the Q3 data set for the first 30 days of the data collection period. This data set spanned  $\sim 90$  days and was the first “typical” quarter of observations as it was 3 months long and did not suffer from excess pointing drift, which necessitated pointing “tweaks” during Q2, or from excessive safe mode events. The effect of the predictor chosen for *Kepler* can be seen in the ramping up of the entropies over the course of each 48-cadence interval ( $\approx 1$  day) as the baseline values for each pixel go stale. There are some occasions when the baseline is anomalous, most likely due to an Argabrightening event when a piece of debris crosses the aperture and reflects light into the barrel of the telescope.<sup>30</sup> Right: Histogram of the differences between the achieved and the theoretical compression rates given the distribution of requantized pixel residuals. The median achieved compression rate is 4.4 bits per pixel, only 0.13 bits per pixel above the theoretical minimum for a perfectly tuned compression table. The achieved compression is within 3% of the theoretical minimum for the chosen compression algorithm.

intervals specified on input. The Huffman Accumulator (HAC) pipeline then aggregated the results across the 84 module outputs and produced diagnostic plots and information so that we could determine whether the default baseline cadence interval produced adequate compression rates. Figure 3 shows that the default baseline interval of 48 Long Cadences is located approximately at the bottom of a very shallow minimum as the overhead associated with storing the uncompressed baseline interval data is traded against better compression for shorter baseline intervals. The predicted compression rate is 4.17 bits per pixel, which is significantly better than the requirement.

The original plan was to update the Huffman table once we were in orbit and had the opportunity to collect a sufficient amount of flight data. However, as the data in Figure 4 shows, this was not necessary. The Calibration (CAL) pipeline examines the original LC pixels that flow through it to be calibrated and calculates the achieved and theoretical entropies for tracking and trending purposes.<sup>29</sup> As can be seen for the third quarter of observations (Q3), the difference between the achieved and theoretical entropies is only 0.13 bits per symbol, or 3%, so in spite of the fact that ETEM does not simulate highly variable, non-solar-like stars, of which there are several thousand to be found in the *Kepler* target catalog, we are achieving a level of compression,  $\sim 4.4$  bits per pixel per cadence, that is very close to the optimum for the compression algorithm selected and implemented in flight and ground software.

#### 4. CHALLENGES TO PHOTOMETRIC PRECISION

Since launch, *Kepler* has provided photometric time series of unprecedented quality and completeness which have in many respects surpassed the expectations of the scientists and engineers who designed and built the spacecraft. There have been some important facts that came to light after *Kepler* returned a significant amount of data. First, while the photometry has been unparalleled in quality, there appears to be excess photometric noise so that the mode of the precision for  $Kp = 12$  G2V stars (solar analogs) on timescales of 6.5 hours is 30 ppm rather than the desired goal of 20 ppm. The *Kepler* team has been studying the data to identify likely causes and has determined that the bulk of the increase in the observed photometric noise is due to stellar variability,<sup>31</sup> which is about twice as high as expected based on *SOHO* observations of the Sun.<sup>3</sup> This is an important discovery, both from the standpoint of understanding our Sun in the context of stars in general, and from that of designing future space transit photometry missions.

The fact that solar-like main-sequence stars exhibit more variability than the Sun on timescales relevant to transit detection does not prevent the *Kepler Mission* from achieving its goals, as increasing the duration of the mission to a period of 8 years will recover the originally scoped sensitivity. However, it does make it that much more important to deal with other challenges facing the attainment of 20 ppm photometric precision, such as various systematic errors and instrumental signatures, which are not always adequately removed from *Kepler* light curves at present.



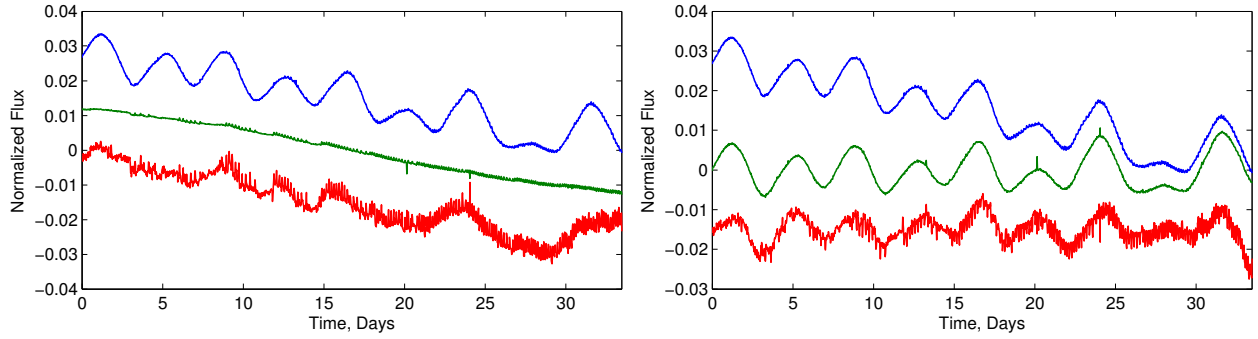


Figure 5. Light curves for one target on the most thermally sensitive CCD channel 2.1 in the *Kepler* focal plane for the Q1 observations. Left: The top curve (blue) is the simple aperture photometry flux time series, the middle (green) curve is the result of a MAP fit with Gaussian priors, and the bottom (red) curve is the result of a robust fit to the same basis vectors. Right: The top (blue) curve is the original flux, the middle (green) curve is the MAP-corrected flux, and the bottom (red) curve is the robust least-squares-corrected flux time series.

#### 4.1 Instrumental Signatures and Systematic Errors

Although *Kepler* is in a very benign, Earth-trailing orbit, there are changes in the thermal state of the telescope as the Sun moves in relation to the telescope. There are also heaters on various spacecraft components used to condition the temperatures of these items within their required bounds. The telescope changes shape in response to the changing distribution of temperatures in the structure, and consequently, the focus of the instrument changes in subtle, but important ways.<sup>17,32</sup> In fact, during the first quarter of science observation we noticed that a heater in the reaction wheel assembly for wheels 3 and 4 was cycling on and off in response to the fact that the box was becoming more shaded during the quarter and the changing temperature of the box was inducing perfectly correlated signatures in the science data at the pixel level. The shape of the telescope was deforming and changing the distance between the primary mirror and the focal plane assembly by  $\sim 1 \mu\text{m}$  in response to this heater. This is a very small change in the state of the telescope, but when the goal is to reach a photometric precision of 20 ppm, it is a significant change. The temperature limits of the thermostat were tightened up so that the cycle time was shortened below one half hour and the temperature swing was significantly reduced.

In the SOC Pipeline, the Pre-search Data Conditioning (PDC) module’s task is to remove systematic errors from the raw flux time series.<sup>33</sup> These include telescopic and spacecraft effects such as pointing errors, focus changes, and changing instrument properties. PDC co-trends each flux time series against ancillary engineering data, such as temperatures of the focal plane and electronics, pointing history, and focus variations, to remove signatures correlated with systematic errors. In general, PDC is doing a good job of correcting systematics for quiet, main-sequence stars, but it is not performing a uniformly good job for all targets, especially for those exhibiting significant intrinsic stellar variability. In some cases PDC removes or grossly distorts intrinsic variations and/or introduces short timescale variations that effectively add noise to the light curves, making it more difficult to detect signatures of transiting planets or to study astrophysics. We also see that in many cases the output light curves from PDC exhibit significant cross correlations, indicating that PDC is not removing all the instrumental signatures and systematics exhibited by the light curves. The basic problem is that PDC is fitting an incomplete model to the data and a maximum likelihood (ML) approach is happy to trade a reduction in bulk *rms* at the cost of introducing excess power on shorter time scales: ML is able to find a happenstance linear combination of basis vectors that match the gross profile of intrinsic stellar variations.

The solution we are implementing is a Bayesian Maximum A Posteriori (MAP) approach whereby we learn about the influence of the systematic errors by examining the behavior of the ensemble of light curves on each CCD readout channel.<sup>34</sup> Figure 5 shows the results of applying a simple version of MAP to Q1 data when the heater on the reaction wheel assembly was active. We performed a singular value decomposition of the most-correlated light curves on this channel to obtain 10 basis vectors that were then robustly fitted to each stellar light curve. We formulated analytical priors from the distributions of the fit coefficients for each term and then applied a MAP fit. The MAP fit was able to “put the brakes” on the least-squares model fit so that highly variable stars were not over-fitted in an attempt to reduce the bulk *rms*, and intrinsic stellar variability was preserved to a high degree while instrumental systematics, such as a long term change in focus that reduced the brightness of the stars, and short term variations in focus caused by the reaction wheel heater, were

effectively removed from the light curves. An empirical version of MAP is being implemented in the SOC PDC pipeline and will be completed and running on *Kepler* data sets by the end of 2011. We expect PDC-MAP to deliver light curves of much higher quality for asteroseismology and astrophysics investigations and for the exoplanet search.

## 5. KEY SCIENCE RESULTS

The *Kepler Mission* has been generating exciting science results right out of the box. The 10 days of continuous observations at the end of commissioning, now dubbed Q0, or Quarter 0, made it clear that detecting transits of giant planets was accomplished easily by eye. Examination of the light curve for HAT-P-7b revealed an  $\approx 100$  ppm drop in brightness  $180^\circ$  out of phase (an occultation) with the 0.6% transits that had been detected from the ground in 2008 before *Kepler* was launched.<sup>35</sup> These small dips are the signature of the atmosphere of the planet heated to  $\approx 2600$  K by its proximity to its star, and were accompanied by phase variations in the light curve indicating that the transport of heat from the dayside to the nightside of the planet is inefficient.<sup>36</sup> This light curve demonstrated unambiguously that *Kepler* can detect transiting planets and furthermore, can detect the weak signatures of Earth-like planets transiting sun-like stars.

The identification of hundreds of potential transiting planets in the initial *Kepler* observations followed quickly on the heels of the occultation of HAT-P-7b. By January of 2010, five planets were announced and hundreds of additional candidates awaited the following observing season for follow up observations and confirmation. The first system of multiple transiting planets was announced in August 2010,<sup>37</sup> consisting of two Saturn-size planets conducting a coupled orbital dance near resonance with 19.24 and 38.91 day periods. These bodies exhibit mutual gravitational interactions in the timing of the transits observed by *Kepler*. These transit timing variations (TTVs) can be used to infer or constrain the masses of the transiting bodies, and can also indicate the presence of non-transiting planets that are perturbing the orbits of the transiting planets. The first unquestionably rocky planet was announced by the *Kepler* team in January of 2011.<sup>38</sup> Radial velocity observations of this planet indicated it has a mass of  $4.5 M_\oplus$ , so with a radius of  $1.4 R_\oplus$  this planet has a density of  $8.8 \text{ g cm}^{-3}$ . Asteroseismology provided a much improved radius estimate for Kepler-10b's star. Kepler-11, an extremely compact system of 6 transiting planets, was discovered with 5 of the 6 planets exhibiting TTVs so that their masses could be estimated directly from the *Kepler* data.<sup>39</sup> The planetary orbits in this system are highly coplanar.

Beyond the individual planet discoveries, the statistics of the *Kepler* results are accumulating to the point that we are learning about the nature and prevalence of planetary formation. Of the 1235 candidates identified by *Kepler*,<sup>1</sup> 408 of these planetary candidates are found in 170 multiple systems, demonstrating the flatness of planetary systems composed of small planets. Analysis of the single transiting candidates and the multiple transiting candidates suggests a relative absence of giant planets in close orbits for the multiple planet systems, and that the false positive rate for the multiple transit systems is significantly lower than that of the single transit planet candidates.<sup>40</sup> The data release of February 2011 has led to a profusion of papers inferring the intrinsic planet frequency from the initial candidate planet catalog. While it is premature to extrapolate these initial findings to the ultimate prize, Earth-size planets orbiting in the habitable zone of solar-like stars, the initial results indicate that the frequency of planets varies inversely with their size, so that smaller planets are more common than giant planets down to at least  $2 R_\oplus$ .<sup>1,41</sup>

As important as *Kepler* is for exoplanets, the asteroseismology community in turn is using this instrument as a near-ideal platform to write new chapters in their field. Initial results from the *Kepler* data suggest that the long-held belief that asteroseismology can reveal the secret inner workings of stars is being confirmed today.<sup>42</sup> The *Kepler* data allow us to quantify global properties of the observed stars, such as radius and age, through observations of normal modes of oscillation. One of the most exciting recent results is the ability of *Kepler* to distinguish between giant stars that are still burning hydrogen in a shell surrounding the accumulating helium core, and those giants that have initiated helium burning in the core.<sup>43</sup> Finally, the *Kepler* data are also being used to study a wide variety of oscillating and pulsating stars, including RR Lyr stars, which can more than double their brightness every 12 hours, as well as cataclysmic variable stars.

## 6. CONCLUSIONS

The *Kepler Mission* is solving an astonishing array of astrophysical mysteries with its unprecedented photometric precision and near continuous observations. The success of *Kepler* is the result of many years of hard work and effort on the part of the thousands of individuals who contributed to the design, implementation, and operation of the instrument and spacecraft and of those who process and analyze the resulting data. Twenty-five years ago the *Kepler* concept was a sci-fi fantasy that only a few individuals, notably the Science Principal Investigator, William J. Borucki, believed in. The protagonists in this

narrative persevered and overcame the many daunting obstacles in the path to the ultimate goal by meeting each challenge as it arose and mustering the resources and talent to solve each problem in turn.

By their nature, transit survey missions rely on collecting time series image data. Such missions must rely on the stability of their observing platforms to achieve the ultra high precision needed to discover small planets. The methodology used by *Kepler* to reduce the data volume to a manageable level, and then to compress the data without loss can be readily applied to future missions to reduce the storage required per sample and to minimize the transmission time required to downlink the original data. It is extremely important to minimize the perturbations to the instrument on timescales from 0.01 to several days. Even in *Kepler*'s benign orbital environment, identifying and removing instrumental systematic errors is a real challenge for such sensitive instruments, and the innovative use of Bayesian methods should be explored.

ESA's PLATO Mission,<sup>44</sup> an ambitious transit survey to discover Earth-size planets around the closest and brightest main-sequence stars may be selected for flight by the end of 2011. It would launch in 2017 or 2018 into an orbit at L2 to survey several large  $> 800 \text{ deg}^2$  fields of view with an instrument composed of  $\sim 30$  small telescopes. Also, at least two transit surveys have been proposed to NASA's current Explorer Mission competition. MIT has proposed the Transiting Exoplanet Survey Satellite (TESS) mission,<sup>45</sup> an optical, all-sky transit survey satellite to be launched into orbit in 2016 for a two-year mission with an instrument composed of multiple telescopes with disjointed fields of view. JPL has also proposed an infrared transit survey mission called ELEKTRA to the Explorer Program.<sup>46</sup> These future missions will face many of the same challenges and will surely benefit from the successes of and the lessons learned from the *Kepler Mission*.

## ACKNOWLEDGMENTS

The authors thanks all the engineers, managers, technicians and support staff at Ball Aerospace Technologies Corporation for delivering such a fine instrument to the world. I thank the *Kepler* Science Team for bringing the nascent science in the data to life and the Science Operations Center team for developing the science processing pipeline. I thank scientific programmers Hema Chandrasekaran for coding the requantization table and Huffman code generation table, and Joe Twicken (SETI Institute) for coding the Huffman histogram generator and accumulator software. I thank Bill Wohler for writing the java code for these components. Funding for the *Kepler Mission* is provided by NASA's Science Mission Directorate.

## REFERENCES

- [1] Borucki, W. J., *et al.*, "Characteristics of planetary candidates observed by *Kepler*. II. Analysis of the first four months of data," *ApJ* 736(1), 19 (2011).
- [2] Jenkins, J. M., D. A. Caldwell, and Borucki, W. J., "Some tests to establish confidence in planets discovered by transit photometry," *ApJ* 564, 495 (2000).
- [3] Jenkins, J. M., "The impact of stellar variability on the detection of transiting terrestrial planets", *ApJ* 575, 493 (2002).
- [4] Young, A. T., "Photometric error analysis. VI. Confirmation of Reiger's theory of scintillation," *AJ* 72, 747–753 (1967).
- [5] Koch, D. G., *et al.*, "*Kepler Mission* design, realized photometric performance, and early science," *ApJL* 713(2), L79–L86 (2010).
- [6] Robinson, L. B., Wei, M. Z., Borucki, W. J., Dunham, E. W., Ford, C. H., and Granados, A. F., "Test of CCD Precision Limits for Differential Photometry," *PASP* 107, 1094 (1995).
- [7] Jenkins, J. M., Borucki, W. J., Dunham, E. W. and McDonald, J. S., "High Precision Photometry with Back-Illuminated CCDs," *ASP Conf. Ser.*, 119, 227–280 (1997).
- [8] Koch, D. G., *et al.*, "CCD photometry tests for a mission to detect Earth-size planets in the extended solar neighborhood," *Proc. SPIE* 4013, 508–519 (2000).
- [9] Jenkins, J. M., Witteborn, F., Koch, D. G., Dunham, E., Borucki, W. J., Updike, T. F., Skinner, M. A., and Jordan, S. P., "Processing CCD images to detect transits of Earth-sized planets: Maximizing sensitivity while achieving reasonable downlink requirements," *Proc. SPIE* 4013, 520–531 (2000).
- [10] Borucki, W., *et al.*, "*Kepler* planet-detection mission: introduction and first results," *Science* 327, 977–980 (2010).
- [11] Ebbets, D., Atcheson, P., Stewart, C., Spuhler, P., "Optical performance of the 100 sq deg FOV telescope for NASA's *Kepler* exoplanet mission," *Proc. SPIE* 8146, 8146-15, this volume (2011).
- [12] Ebbets, D., Argabright, V., VanCleve, J., Caldwell, D. "In-flight performance of the 96 Mpx focal plane for NASA's *Kepler* mission," *Proc. SPIE* 8146, 8146-16, this volume (2011).

- [13] Batalha, N. M., *et al.*, “Selection, prioritization, and characteristics of *Kepler* target stars,” *ApJL* 713(2), L109-L114 (2010).
- [14] Brown, T. M., Latham, D. W., Everett, M. E., and Esquerdo, G. A., “*Kepler* input catalog: Photometric calibration and stellar classification,” *AJ*, in press, arXiv:1102.0342 (2011).
- [15] Bryson, S. T., *et al.*, “Selecting pixels for *Kepler* downlink,” *Proc. SPIE* 7740, 77401D (2010).
- [16] Bryson, S. T., *et al.*, “The *Kepler* pixel response function,” *ApJL* 713(2), L97–L102 (2010).
- [17] Jenkins, J. M., *et al.*, “Overview of the *Kepler* science processing pipeline,” *ApJL* 713(2), L87-L91 (2010).
- [18] Quintana, E. M., *et al.*, “Pixel-level calibration in the *Kepler* Science operations center pipeline,” *Proc. SPIE* 7740, 77401X (2010).
- [19] Tenenbaum, P., and Jenkins, J. M., “Focal plane geometry characterization of the *Kepler Mission*” *Proc. SPIE* 7740, 77401C (2010).
- [20] Haas, M., *et al.*, “*Kepler* science operations,” *ApJL* 713(2), L115–L119 (2010).
- [21] Chandrasekaran, H., *et al.*, “Semi-weekly monitoring of the performance and attitude of *Kepler* using a sparse set of targets,” *Proc. SPIE* 7740, 77401B (2010).
- [22] Van Cleve, J., and Caldwell, D. A., [*Kepler Instrument Handbook*, KSCI 19033-001], NASA Ames Research Center, Moffett Field, CA (2009).
- [23] Papoulis, A., [*Probability, Random Variables, and Stochastic Processes*], McGraw Hill, New York (1984).
- [24] Huffman, D. A., “A Method for the Construction of Minimum-Redundancy Codes,” *Proc. Inst. Radio. Eng.* 40, 1098-1101 (1952).
- [25] Larmore, L. L., and Hirschberg, D. S., “A fast algorithm for optimal length-limited Huffman codes,” *J. Assoc. Comput. Mach.* 37(3), 464–473 (1990).
- [26] Katajainen, J., Moffat, A., and Turpin, A., “A fast and space-economical algorithm for length-limited coding,” *Proc. of the 6th Int. Symp. on Algorithms and Computation*, 1004, 12 – 21 (1995).
- [27] Jenkins, J. M., Peters, D. J., and Murphy, D. W., “An efficient end-to-end model for the *Kepler* photometer,” *Proc. SPIE* 5497, 202-212 (2004).
- [28] Bryson, S. T., Jenkins, J. M., Peters, D. J., Tenenbaum, P. P., Klaus, T. C., Gunter, J. P., Cote, M. T., and Caldwell, D. A., “The *Kepler* end-to-end model: creating high-fidelity simulations to test *Kepler* ground processing,” *Proc. SPIE* 7738, 773808 (2010).
- [29] Li, J., *et al.*, “Photometer performance assessment in *Kepler* science data processing,” *Proc. SPIE* 7740, 77401T (2010).
- [30] Witteborn, F. C., Van Cleve, J., Borucki, W. J., Argabright, V. S., and Hascall, P., “Debris sightings in the *Kepler* field,” *Proc. SPIE* 8151, 8151-42 (2011).
- [31] Gilliland, R. L., *et al.*, “*Kepler Mission* stellar and instrument noise properties,” *ApJ*, in press, arXiv:1107.5207 (2011).
- [32] Jenkins, J. M., *et al.*, “Initial characteristics of *Kepler* long cadence data for detecting transiting planets,” *ApJL* 713(2), L120-L125 (2010).
- [33] Twicken, J. T., Chandrasekaran, H., Jenkins, J. M., Gunter, J. P., Girouard, F., Klaus, T. C., “Presearch data conditioning in the *Kepler* science operations center pipeline,” *Proc. SPIE* 7740, 77401U (2010).
- [34] Jenkins, J. M., Smith, J. C., Tenenbaum, P., Twicken, J. T., “Planet detection: the *Kepler Mission*,” [*Advances in Machine Learning and Data Mining for Astronomy*], Chapman and Hall/CRC Press, in press (2011).
- [35] Pál, A., *et al.*, “HAT-P-7b: An extremely hot massive planet transiting a bright star in the *Kepler* field,” *ApJ* 680(2), 1450-1456 (2008).
- [36] Borucki, W. J., *et al.*, “*Kepler*’s optical phase curve of the exoplanet HAT-P-7b,” *Science* 325, 709 (2009).
- [37] Holman, M. *et al.*, “*Kepler*-9: A system of multiple planets transiting a sun-Like star, confirmed by timing variations,” *Science* 330, 51 (2010).
- [38] Batalha, N. M., *et al.*, “*Kepler*’s first rocky planet: *Kepler*-10b,” *ApJ* 729, 27 (2011).
- [39] Lissauer, J. J., *et al.*, “A closely packed system of low-mass, low-density planets transiting *Kepler*-11,” *Nature* 470, 53–58.
- [40] Latham, D., *et al.*, “A first comparison of *Kepler* planet candidates in single and multiple systems,” *ApJL* 732, L24 (2011).

- [41] Howard, A. W., *et al.*, “Planet occurrence within 0.25 AU of solar-type stars from *Kepler*” ApJ, arXiv:1103.2541 (2011).
- [42] Gilliland, R. L., *et al.*, “*Kepler* asteroseismology program: Introduction and first results,” PASP 122, 131–143 (2010).
- [43] Bedding, T. R., *et al.*, “Gravity modes as a way to distinguish between hydrogen- and helium-burning red giant stars,” Nature 471, 608–611 (2011).
- [44] Claudi, R., “A new opportunity from space: PLATO mission,” Ap&SS 328, 319–323 (2010).
- [45] Ricker, G., *et al.*, “Transiting Exoplanet Survey Satellite (TESS),” BAAS 42, 459 (2010).
- [46] Beichman, C. and the ELEKTRA Team, “The ELEKTRA explorer mission to find transiting earth like planets,” presented at the Strange New Worlds Conference held in Flagstaff AZ (2011).