arXiv:1103.2541v1 [astro-ph.EP] 13 Mar 2011

PLANET OCCURRENCE WITHIN 0.25 AU OF SOLAR-TYPE STARS FROM *KEPLER*[†]

Andrew W. Howard[1,*], Geoffrey W. Marcy[1], Stephen T. Bryson[2], Jon M. Jenkins[3], Jason F. Rowe[2],
Natalie M. Batalha[4], William J. Borucki[2], David G. Koch[2], Edward W. Dunham[5], Thomas N. Gautier III[6],
Jeffrey Van Cleve[3], William D. Cochran[7], David W. Latham[8], Jack J. Lissauer[2], Guillermo Torres[8],
Timothy M. Brown[9], Ronald L. Gilliland[10], Lars A. Buchhave[11], Douglas A. Caldwell[3],
Jørgen Christensen-Dalsgaard[12,13], David Ciardi[14], Francois Fressin[8], Michael R. Haas[2], Steve B. Howell[15],
Hans Kjeldsen[12], Sara Seager[16], Leslie Rogers[16], Dimitar D. Sasselov[8], Jason H. Steffen[17], Gibor S. Basri[1],
David Charbonneau[8], Jessie Christiansen[2], Bruce Clarke[2], Andrea Dupree[8], Daniel C. Fabrycky[18],
Debra A. Fischer[19], Eric B. Ford[20], Jonathan J. Fortney[18], Jill Tarter[3], Forrest R. Girouard[21],
Matthew J. Holman[8], John Asher Johnson[22], Todd C. Klaus[21], Pavel Machalek[3], Althea V. Moorhead[20],
Robert C. Morehead[20], Darin Ragozzine[8], Peter Tenenbaum[3], Joseph D. Twicken[3], Samuel N. Quinn[8],
Howard Isaacson[1], Avi Shporer[9,23], Philip W. Lucas[24], Lucianne M. Walkowicz[1], William F. Welsh[25],
Alan Boss[26], Edna Devore[3], Alan Gould[27], Jeffrey C. Smith[3], Robert L. Morris[3], Andrej Prsa[28],
Timothy D. Morton[21]
Submitted to ApJ — 2011/03/13

ABSTRACT

We report the distribution of planets as a function of planet radius, orbital period, and stellar effective temperature for orbital periods less than 50 days around Solar-type (GK) stars. These results are based on the 1,235 planets (formally "planet candidates") from the *Kepler* mission that include a nearly complete set of detected planets as small as 2 $R_\oplus$. For each of the 156,000 target stars we assess the detectability of planets as a function of planet radius, $R_{\rm p}$, and orbital period, P, using a measure of the detection efficiency for each star. We also correct for the geometric probability of transit, $R_\star/a$. We consider first *Kepler* target stars within the "solar subset" having $T_{\rm eff}$ = 4100–6100 K, $\log g$ = 4.0–4.9, and Kepler magnitude Kp < 15 mag, i.e. bright, main sequence GK stars. We include only those stars having photometric noise low enough to permit detection of planets down to 2 $R_\oplus$. We count planets in small domains of $R_{\rm p}$ and P and divide by the included target stars to calculate planet occurrence in each domain. The resulting occurrence of planets varies by more than three orders of magnitude in the radius-orbital period plane and increases substantially down to the smallest radius (2 $R_\oplus$) and out to the longest orbital period (50 days, ∼0.25 AU) in our study. For P < 50 days, the distribution of planet radii is given by a power law, $df/d\log R = k_R R^\alpha$ with $k_R = 2.9^{+0.5}_{-0.4}$, α = −1.92 ± 0.11, and $R = R_{\rm p}/R_\oplus$. This rapid increase in planet occurrence with decreasing planet size agrees with the prediction of core-accretion formation, but disagrees with population synthesis models that predict a desert at super-Earth and Neptune sizes for close-in orbits. Planets with orbital periods shorter than 2 days are extremely rare; for $R_{\rm p}$ > 2 $R_\oplus$ we measure an occurrence of less than 0.001 planets per star. For all planets with orbital periods less than 50 days, we measure occurrence of 0.130 ± 0.008, 0.023 ± 0.003, and 0.013 ± 0.002 planets per star for planets with radii 2–4, 4–8, and 8–32 $R_\oplus$, in agreement with Doppler surveys. We fit occurrence as a function of P to a power law model with an exponential cutoff below a critical period P_0. For smaller planets, P_0 has larger values, suggesting that the "parking distance" for migrating planets moves outward with decreasing planet size. We also measured planet occurrence over a broader stellar $T_{\rm eff}$ range of 3600–7100 K, spanning M0 to F2 dwarfs. Over this range, the occurrence of 2–4 $R_\oplus$ planets in the *Kepler* field linearly increases with decreasing $T_{\rm eff}$, making these small planets seven times more abundant around cool stars (3600–4100 K) than the hottest stars in our sample (6600–7100 K).

Subject headings: planetary systems, stars: statistics — techniques: photometry

[1] University of California, Berkeley, CA 94720
[2] NASA Ames Research Center, Moffett Field, CA 94035
[3] SETI Institute/NASA Ames Research Center, Moffett Field, CA 94035
[4] San Jose State University, San Jose, CA 95192
[5] Lowell Observatory, Flagstaff, AZ 86001
[6] Jet Propulsion Laboratory/Caltech, Pasadena, CA 91109
[7] University of Texas, Austin, TX 78712
[8] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138
[9] Las Cumbres Observatory Global Telescope, Goleta, CA 93117
[10] Space Telescope Science Institute, Baltimore, MD 21218
[11] Niels Bohr Institute, Copenhagen University, Denmark
[12] Aarhus University, DK-8000 Aarhus C, Denmark
[13] High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO 80307
[14] NASA Exoplanet Science Institute/Caltech, Pasadena, CA 91125
[15] National Optical Astronomy Observatory, Tucson, AZ 85719
[16] Massachusetts Institute of Technology, Cambridge, MA, 02139
[17] Fermilab Center for Particle Astrophysics, Batavia, IL 60510
[18] University of California, Santa Cruz, CA 95064
[19] Yale University, New Haven, CT 06510
[20] University of Florida, Gainesville, FL 32611
[21] Orbital Sciences Corp., NASA Ames Research Center, Moffett Field, CA 94035

[22] California Institute of Technology, Pasadena, CA 91109

[23] Department of Physics, Broida Hall, University of California, Santa Barbara, CA 93106

[24] Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, England

[25] San Diego State University, San Diego, CA 92182

[26] Carnegie Institution of Washington, Dept. of Terrestrial Magnetism, Washington, DC 20015

[27] Lawrence Hall of Science, Berkeley, CA 94720

[28] Villanova University, Dept. of Astronomy and Astrophysics, 800 E Lancaster Ave, Villanova, PA 19085

[†] Based in part on observations obtained at the W. M. Keck Observatory, which is operated by the University of California and the California Institute of Technology.

[*] To whom correspondence should be addressed. E-mail: howard@astro.berkeley.edu

1. INTRODUCTION

The dominant theory for the formation of planets within 20 AU involves the collisions and sticking of planetesimals having a rock and ice composition, growing to Earth-size and beyond. The presence of gas in the protoplanetary disk allows gravitational accretion of hydrogen, helium and other volatiles, with accretion rates depending on gas density and temperature, and hence on location within the disk and its stage of evolution. The relevant processes, including inward migration, have been simulated numerically both for individual planet growth and for entire populations of planets (Ida & Lin 2004, 2008b; Mordasini et al. 2009a; Schlaufman et al. 2010; Ida & Lin 2010; Alibert et al. 2011).

The simulations suggest that most planets form near or beyond the ice line. When they reach a critical mass of several Earth-masses ($M_\oplus$) the planets either rapidly spiral inward to the host star because of the onset of Type II migration or undergo runaway gas accretion and become massive gas-giants, thus producing a "planet desert" (Ida & Lin 2008a). The predicted desert resides in the mass range $\sim$1–20 $M_\oplus$ orbiting inside of $\sim$1 AU, with details that vary with assumed behavior of inward planet migration (Ida & Lin 2008b, 2010; Alibert et al. 2011; Schlaufman et al. 2009). Another prediction is that the distribution of planets in the mass/orbital distance plane is fairly uniform for masses above the planet desert ($\gtrsim$ 20 $M_\oplus$) and inside of $\sim$0.25 AU (periods less than 50 days). The majority of the planets in these models reside near or beyond the ice line at $\sim$2 AU (well outside of the $P < 50$ days domains analyzed here). The mass distribution for these distant planets rises toward super-Earth and Earth-mass (Ida & Lin 2008b; Mordasini et al. 2009b; Alibert et al. 2011). These patterns of planet occurrence in the two-parameter space defined by planet masses and orbital periods can be directly tested with observations of a statistically large sample of planets orbiting within 1 AU of their host stars.

Two early observational tests of the planet-formation simulations have emerged. Using RV-detected planets, Howard et al. (2010) measured planet occurrence for close-in planets ($P < 50$ days) with masses that span nearly three orders of magnitude—super-Earths to Jupiters ($M_\mathrm{p} \sin i = $ 3–1000 $M_\oplus$). This Eta-Earth Survey focused on 166 G and K dwarfs on the main sequence. The survey showed an increasing occurrence, f, of planets with decreasing mass, M, from 1000 to 3 $M_\oplus$. A power law fit to the observed distribution of planet mass gave $\mathrm{d}f/\mathrm{d}\log M = 0.39 M^{-0.48}$. Remarkably, the survey revealed a high occurrence of planets in the period range $P = $ 10–50 days and mass range $M_\mathrm{p} \sin i = $ 4–10 $M_\oplus$, precisely within the predicted planet desert. Planets with $M_\mathrm{p} \sin i = $ 10–100 $M_\oplus$ and $P < 20$ days were found to be quite rare. Thus, the predicted desert was found to be full of planets and the predicted uniform mass distribution for close-in planets above the desert was found to be rising with smaller mass, not flat. These discrepancies suggest that current population synthesis models of planet formation around solar-type stars are somehow failing to explain the distribution of low-mass planets around solar-type stars.

Accounting for completeness, Howard et al. (2010) found a planet occurrence of $15^{+5}_{-4}\%$ for planets with $M_\mathrm{p} \sin i = $ 3–30 $M_\oplus$ and $P < 50$ d around main sequence G and K stars. In contrast, Mayor et al. have asserted a substantially higher planet occurrence of 30% ± 10% (Mayor et al. 2009) or higher with a careful statistical study still in progress. Thus, there may be observational discrepancies in planet occurrence which we expect to be resolved soon. Still, there is qualitative agreement between Howard et al. (2010) and Mayor et al. (2009) that the predicted paucity of planets of mass $\sim$1–30 $M_\oplus$ within 1 AU is not observed, as that close-in domain is, in fact, rich with small planets. The planet candidates from *Kepler*, along with a careful assessment of both false positive rates and completeness, can add a key independent measure of the occurrence of small planets to compare with the Eta-Earth Survey and Mayor et al. Formally these objects are "planet candidates" as a small percentage will turn out to be false positive detections; we often refer to them as "planets" below.

The observed occurrence of small planets orbiting close-in matches continuously with the similar analysis by Cumming et al. (2008) who measured 10.5% of Solar-type stars hosting a gas-giant planet ($M_\mathrm{p} \sin i = $ 100–3000 $M_\oplus$, $P = $ 2–2000 days), for which planet occurrence varies as $\mathrm{d}f \propto M^{-0.31\pm0.2} P^{0.26\pm0.1} \, \mathrm{d}\log M \, \mathrm{d}\log P$. Thus, the occurrence of giant planets orbiting in 0.5–3 AU seems to attach smoothly to the occurrence of planets down to 3 $M_\oplus$ orbiting within 0.25 AU. This suggests that the formation and accretion processes are continuous in that domain of planet mass and orbital distance, or that the admixture of relevant processes varies continuously from 1000 $M_\oplus$ down to 3 $M_\oplus$.

Planet formation theory must also account for remarkable orbital properties of exoplanets. The orbital eccentricities span the range $e = $ 0–0.93 and the close-in "hot Jupiters" show a wide distribution of alignments (or misalignments) with the equatorial plane of the host star (e.g., Johnson et al. 2009; Winn et al. 2010, 2011; Triaud et al. 2010; Morton & Johnson 2010). Thus, standard planet formation theory probably requires additional planet-planet gravitational interactions to explain these non-circular and non-coplanar orbits (e.g. Chatterjee et al. 2010; Wu & Lithwick 2011; Nagasawa et al. 2008).

The distribution of planets in the mass/orbital-period plane reveals important clues about planet formation and migration. Here we carry out an analysis of the epochal *Kepler* results for transiting planet candidates from Borucki et al. (2011) with a careful treatment of the completeness. We focus attention on the planets with orbital periods less than 50 days to match the period range that RV surveys are most sensitive to. The goals are to measure the occurrence distribution of close-in planets, to independently test planet population synthesis models, and to check the Doppler RV results of Howard et al. (2010). While none of the planets or stars are in common between *Kepler* and RV surveys, we will combine the mass distribution (from RV) and the radius distribution (from *Kepler*) to constrain the bulk densities of the *types* of planets they have in common. Planet formation models predict great diversity in the interior structures of planets having Earth-mass to Saturn-mass, caused by the various admixtures of rock, water-ice, and H & He gas. Here we attempt to statistically assess planet radii

and masses to arrive for the first time at the density distribution of planets within 0.25 AU of their host stars.

2. SELECTION OF KEPLER TARGET STARS AND PLANET CANDIDATES

We seek to determine the occurrence of planets as a function of orbital period, planet radius (from *Kepler*) and planet mass (from Doppler searches). Measuring occurrence using either Doppler or transit techniques suffers from detection efficiency that is a function of both the properties of the planet (radius, orbital period) and of the individual stars (notably noise from stellar activity). Thus the effective stellar sample from which occurrence may be measured is itself a function of planet properties and the quality of the data for each target star. A key element of this paper is that *only a subset of the target stars are amenable to the detection of planets having a certain radius and period.*

To overcome this challenge posed by planet detection completeness, we construct a two-dimensional space of orbital period and planet radius (or mass). We divide this space into small domains of specified increments in period and planet radius (or mass) and carefully determine the subset of target stars for which the detection of planets in that small domain has high efficiency. In that way, each domain of orbital period and planet size (or mass) has its own subsample of target stars that are selected *a priori*, within which the detected planets can be counted and compared to that number of stars. This treatment of detection completeness for each target star was successfully adopted by Howard et al. (2010) in the assessment of planet occurrence as a function of orbital period and planet mass ($M_{\rm p} \sin i$) from Doppler surveys. Here we carry out a similar analysis of occurrence of planets from the *Kepler* survey in a two-dimensional space of orbital period and planet radius.

2.1. *Winnowing the Kepler Target Stars for High Planet Detectability*

To measure planet occurrence we compare the number of detected planets having some set of properties (radii, orbital periods, etc.) to the set of stars from which planets with those properties could have been reliably detected. Errors in either the number of planets detected or the number of stars surveyed corrupt the planet occurrence measurement. We adopt the philosophy that it is preferable to suffer higher Poisson errors from considering fewer planets and stars than the difficult-to-quantify systematic errors caused by studying a larger number of planets and stars with more poorly determined detection completeness.

We begin our winnowing of target stars with the *Kepler* Input Catalog (KIC; Brown et al. 2011; Kepler Mission Team 2009). In this paper we include only planet candidates found in three data segments ("Quarters") labeled Q0, Q1, and Q2, for which all photometry is published (Borucki et al. 2011). Q0 was data commissioning (2–11 May 2009), Q1 includes data from 13 May to 15 June 2009, and Q2 includes data from 15 June to 17 September 2009. The segments had durations of 9.7, 33.5, and 93 days, respectively. *Kepler* achieved a duty cycle of greater than 90%, which almost completely eliminated window function effects (von Braun et al. 2009). A total of 156,097 long cadence

targets (30 min integrations) were observed in Q1 and 166,247 targets were observed in Q2, with the targets in Q2 being nearly a superset of those in Q1. In this paper we consider only the "exoplanet target stars" of which there were 153,196 observed during Q2, and are used for the statistics presented here (Batalha et al. 2010). (The remaining *Kepler* targets in Q2 were evolved stars, not suitable for sensitive planet detection.) The few percent changes in the planet-search target stars are not significant here as Q2 data dominate the planet detectability. The KIC contains stellar $T_{\rm eff}$ and radii ($R_\star$) that are based on four visible-light magnitudes (g, r, i, z) and a fifth, *D51*, calibrated with model atmospheres, and *JHK* IR magnitudes (Brown et al. 2011).

The photometric calibrations yield $T_{\rm eff}$ reliable to ± 135 K (rms) and surface gravity $\log g$ reliable to ± 0.25 dex (rms), based on a comparison of KIC values to results of high resolution spectra obtained with the Keck I telescope and LTE analysis (Brown et al. 2011). Stellar radii are estimated from $T_{\rm eff}$ and $\log g$ and carry an uncertainty of 0.13 dex, i.e. 35% rms (Brown et al. 2011). There is a concern that values of $\log g$ for subgiants are systematically overestimated, leading to stellar radii that are smaller than their true radii perhaps by as much as a factor of two. One should be concerned that a magnitude-limited survey such as *Kepler* may favor slightly evolved stars, implying systematic underestimates of stellar radii, an effect worth considering at the interpretation stage of this work. The quoted planet radii may be too small by as much as a factor of two for evolved stars. We adopt these KIC values for stellar $T_{\rm eff}$ and $R_\star$ from the KIC and their associated uncertainties, following Borucki et al. (2011). The stellar metallicities are poorly known. The KIC is available on the Multi-Mission Archive at the Space Telescope Science Institute (MAST) website[31].

In this paper, we primarily consider *Kepler* target stars having properties in the core of the *Kepler* mission, namely bright solar-type main sequence stars. Specifically, we consider only Kepler target stars within this domain of the H-R diagram: $T_{\rm eff} = 4100$–6100 K, $\log g = 4.0$–4.9, and Kepler magnitude Kp < 15 mag (Table 1). These parameters select for the brightest half of the GK-type target stars (the other half being fainter, Kp > 15 mag), as shown in Figure 1. The goal is to limit our study to main sequence GK stars well characterized in the KIC (Brown et al. 2011) and to provide a stellar sample that is a close match to that of Howard et al. (2010), offering an opportunity for a comparison of the radii and masses from the two surveys. The brightness limit of Kp < 15 promotes high photometric signal-to-noise ratios, needed to detect the smaller planets. These three criteria in $T_{\rm eff}$, $\log g$, and Kp seem, at first glance, to be quite modest, representing the core target stars in the *Kepler* mission. *Yet these three stellar criteria yield a subsample of only 58,041 target stars, roughly one third of the total Kepler sample.* In this study, we consider only this subset of *Kepler* stars and the associated planet candidates detected among them.

2.2. *Winnowing Kepler Target Stars by Detectable Planet Radius and Period*

TABLE 1
PROPERTIES OF STELLAR AND PLANETARY SAMPLES

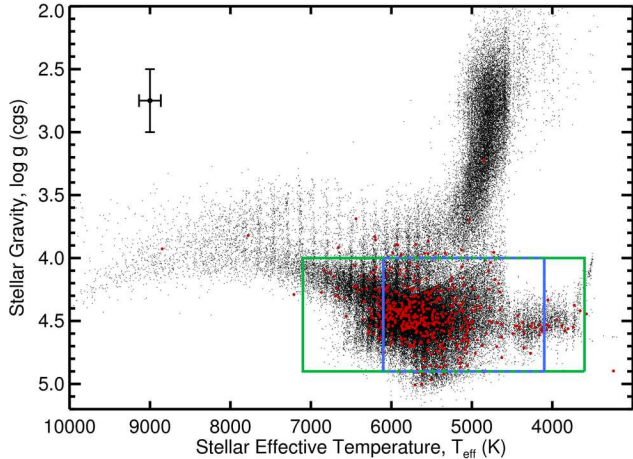| Parameter	Value
Stellar effective temperature, $T_{\rm eff}$	4100–6100 K
Stellar gravity, $\log g$ (cgs)	4.0–4.9
Kepler magnitude, Kp	< 15
Number of stars, $n_\star$	58,041
Orbital period, P	< 50 days
Planet radius, $R_{\rm p}$	2–32 $R_\oplus$
Detection threshold, SNR (90 days)	> 10
Number of planet candidates, $n_{\rm pl}$	438

FIG. 1.— *Kepler* target stars (small black dots) and *Kepler* stars with planet candidates (red dots) plotted as a function of $T_{\rm eff}$ and $\log g$ from the KIC. Only bright stars (Kp < 15) are shown and considered in this study. The inner blue rectangle marks the "solar subset" ($T_{\rm eff} = 4100$–6100 K and $\log g = 4.0$–4.9) of main sequence G and K stars considered for most of this study. This domain contains 58,041 stars with 438 planet candidates. In Section 4 we consider planet occurrence as a function of $T_{\rm eff}$. For that analysis we consider a broader range of $T_{\rm eff} = 3600$–7100 K (green outer rectangle). The error bars in the upper left show the typical uncertainties of 135 K in $T_{\rm eff}$ and 0.25 dex in $\log g$.

We further restrict the *Kepler* stellar sample by including only those stars with high enough photometric quality to permit detection of planets of a specified radius and orbital period. To begin, we consider differential domains in the two-dimensional space of planet radius and orbital period. For each differential domain, only a subset of the *Kepler* target stars have sufficient photometric quality to permit detection of such a planet. In effect, the survey for such specific planets is carried out *only* among those stars having photometric quality so high that the transit signals stand out easily (literally by eye). For photometric quality we adopt the metric of the signal-to-noise ratio (SNR) of the transit signal integrated over a 90 day photometric time series. We define SNR to be the transit depth divided by the uncertainty in that depth due to photometric noise (to be defined quantitatively below).

We set a threshold, SNR > 10, which is higher than that (SNR > 7.0) adopted by Borucki et al. (2011), lending our study an even higher standard of detection. Thus, we restrict our sample of stars so strongly that planets of a specified radius and orbital period are rarely, if ever, missed by the "Transiting Planet Search" (TPS;

Jenkins et al. 2010c) pipeline. Moreover, we base our SNR criterion on just a single 90 day quarter of *Kepler* photometry. This conservatively demands that the photometric pipeline detect transits only during a single pointing of the telescope. (The CCD pixels that a particular star falls on change quarterly as *Kepler* is rolled by 90 degrees to maintain solar illumination.) As noted in Borucki et al. (2011), the photometric pipeline does not yet have the capability to stitch together multiple quarters of photometry and search for transits. In contrast, the SNR quoted in Borucki et al. (2011) was based on the totality of photometry, Q0–Q5 (approximately one year in duration). Thus we are setting a threshold that is considerably more stringent than in Borucki et al. (2011), i.e. including target stars of the quietest photometric behavior. The goal, described in more detail below, is to establish a subset of *Kepler* target stars for which the detection efficiency of planets (of specified radius and orbital period) is close to 100%.

Finally, we restrict our study to orbital periods under 50 days. All criteria by which *Kepler* stars are retained in our study are given in Table 1. As demonstrated below, these restrictions on SNR > 10 and on orbital period ($P < 50$ days) yield a final subsample of *Kepler* targets for which very few planet candidates will be missed by the current *Kepler* photometric pipeline as the transit signals both overwhelm the noise and repeat multiple times (for $P < 50$ days).

We explored the adoption of two measures of photometric SNR for each *Kepler* star, one taken directly from Borucki et al. (2011) and the other using the so-called Combined Differential Photometric Precision ($\sigma_{\rm CDPP}$) which is the empirical RMS noise in bins of a specified time interval, coming from the *Kepler* pipeline. Actually, Borucki et al. (2011) derived their SNR values from $\sigma_{\rm CDPP}$, integrated over all transits in Q0–Q5. We employed the measured $\sigma_{\rm CDPP}$ for time intervals of 3 hr and compared the resulting SNR from Borucki et al. (2011) for transits to those we computed from the basic $\sigma_{\rm CDPP}$. These values agreed well (understandably, accounting for the use of a total SNR from all five quarters in Borucki et al. (2011)). Thus, we adopted the basic 3 hr $\sigma_{\rm CDPP}$ for each target star as the origin of our noise measure.

Each Kepler target star has its own measured RMS noise level, $\sigma_{\rm CDPP}$. Typical 3 hr $\sigma_{\rm CDPP}$ values are 30–300 parts per million (ppm), as shown in Figure 1 of Jenkins et al. (2010b), albeit for 6 hr time bins. Clearly, the photometrically noisiest target stars are less amenable to the detection of small planets, which we treat below. The noise has three sources. One is simply Poisson errors from the finite number of photons received, dependent on the star's brightness, causing fainter stars to have higher $\sigma_{\rm CDPP}$. This photon-limited photometric noise is represented by the lower envelope of the noise as a function of magnitude in Figure 1 of Jenkins et al. (2010b). A second noise source stems from stellar surface physics including spots, convective overshoot and turbulence (granulation), acoustic p-modes, and magnetic effects arising from plage regions and reconnection events. A third noise source stems from excess image motion in Q0, Q1, and Q2 stemming from use of variable guide stars that have now been dropped. In Q2 the presence of bulk drift corrected by four re-

pointings of the bore sight, plus a safe mode followed by an unusually large thermal recovery also contributed. The measured $\sigma_{\rm CDPP}$ accounts for all such sources, as well as any unmentioned since it is an empirical measure.

Using $\sigma_{\rm CDPP}$ for each target star, we define SNR integrated over all transits as,

$$\text{SNR} = \frac{\delta}{\sigma_{\rm CDPP}} \sqrt{\frac{n_{\rm tr} \cdot t_{\rm dur}}{3 \, \text{hr}}}. \qquad (1)$$

Here $\delta = R_{\rm p}^2/R_\star^2$ is the photometric depth of a central transit of a planet of radius $R_{\rm p}$ transiting a star of radius $R_\star$, $n_{\rm tr}$ is the number of transits observed in a 90 day quarter, $t_{\rm dur}$ is the transit duration, and the factor of 3 hr accounts for the duration over which $\sigma_{\rm CDPP}$ was measured. We include only those stars yielding SNR > 10, for a given specified transit depth and orbital period. The threshold imposes such a stringent selection of target stars that few planets are missed by the *Kepler* Transiting Planet Search (TPS) pipeline. Our planet occurrence analysis below assumes that (nearly) all planets with $R_{\rm p} > 2R_\oplus$ that meet the above SNR criteria have been detected by the *Kepler* pipeline and are included in Borucki et al. (2011). The *Kepler* team is currently engaged in a considerable study of the completeness of the *Kepler* pipeline by injecting simulated transit signals into pipeline at the CCD pixel level and measuring the recovery rate of those signals as a function of SNR and other parameters. In advance of the results of this major numerical experiment, we demonstrate detection completeness of SNR > 10 signals in two ways.

First, Figure 2 shows the SNR of detected transits as a function of *Kepler* Object of Interest (KOI) number. The *Kepler* photometry and TPS pipeline detects planet candidates over the course of months as data arrive. There is a learning curve involved with this process, as both software matures and human intervention is tuned (Rowe et al. 2010). As a result, the obvious (high SNR) planet candidates are issued low KOI numbers as they are detected early in the mission. The shallower transits, relative to noise, are identified later as they require more data, and are issued larger KOI numbers. Thus KOI number is a rough proxy for the time required to accumulate enough photometry to identify the planet candidate. Among the KOIs 1050–1600, much less vetting was done, and indeed we rejected five planet candidates (KOIs 1187, 1227, 1387, 1391, and 1465) reported in Borucki et al. (2011) based on both V-shaped light curves and at least one other property indicating a likely eclipsing binary.

Figure 2 shows that the early KOIs, 1–1000, had a wide range of SNR values spanning 7–1000, as the first transit signals had a variety of depths. KOIs 400–1000 correspond to pipeline detections of transit planet candidates around target stars as faint as 15th mag and fainter. The more recent transit identifications of KOIs 1000–1600 exhibit far fewer transits with SNR > 20 and about half of these new KOIs have SNR < 10, below our threshold for inclusion. Apparently most newly identified KOIs have SNR < 20, and few planets remain to be found with P < 50 days and SNR > 20. *Figure 2 suggests that the great majority of planet candidates with P < 50 days and SNR > 10 have already been identified*

TABLE 2
Properties of Planet Candidates in Figure 3

KOI	Kp (mag)	$R_\star$ ($R_\odot$)	$R_{\rm p}$ ($R_\oplus$)	P (days)	SNR (Q0–Q5)	SNR (90 days)
223.02	14.7	0.74	2.40	41.0	25	12.3
542.01	14.4	1.13	2.70	41.9	21	11.2
592.01	14.3	1.08	2.70	39.8	19	9.7
711.01	14.0	1.00	2.74	44.7	34	25.3

by the Kepler *pipeline.* This apparent asymptotic success in the detection of SNR > 10 transits is enabled by our orbital period limit of 50 days which is considerably less than the duration of a quarter (90 days). The current *Kepler* pipeline for identifying transits within a single 90 day quarter is more robust than the multi-quarter transit search. For such short periods, at least two transits typically occur within one quarter. Moreover, when such planet candidates appear during another quarter, the short period planets are quickly confirmed. We suspect that for periods greater than 90 days, many more planet candidates are yet to be identified by *Kepler*. Thus, this study restricts itself to P < 50 days in part because of the demonstrated completeness of detection for such short periods.

We examined the light curves themselves for a second demonstration of nearly complete detection efficiency of planet candidates with P < 50 days, $R_{\rm p}$ > 2 $R_\oplus$, and SNR > 10. Figure 3 shows four representative light curves of planet candidates whose properties are listed in Table 2. All four have small radii of 2–3 $R_\oplus$ and "long" periods of 30–50 days, the most difficult domain for planet detection in this study (the lower right corner of Figure 4, discussed below). The SNR values are near the threshold value of ~10; in fact, one planet candidate (KOI 592.01) has a SNR of 9.7 and is therefore conservatively excluded from this study. *The four light curves show how clearly such transits stand out, indicating the high detection completeness of planets down to 2 $R_\oplus$ and P < 50 days for the SNR > 10 threshold we adopted.*

2.3. *Identifying Kepler Planet Candidates*

We adopt the *Kepler* planet candidates and their orbital periods and planet radii from Table 2 of Borucki et al. (2011), with two exceptions. First, we exclude the five KOIs noted above that are likely to be false positives. Second, we exclude KOIs that orbit "unclassified" KIC stars (identified with "$T_{\rm eff}$ Flag" = 1 in Table 1 of Borucki et al. (2011)). We measure planet occurrence only around stars with well defined stellar parameters from the KIC.

To summarize the Borucki et al. (2011) results, photometry at roughly 100 ppm levels in 29.4 minute integrations allows detection of repeated, brief drops in stellar brightness caused by planet transits across the star. The technical specifics of the instrument, photometry, and transit detection are described in Borucki et al. (2010a); Koch et al. (2010b); Jenkins et al. (2010b,c); Caldwell et al. (2010). We begin the identification of planet candidates based on those revealed in public *Kepler* photometric data (Q0–Q2). This data release contains 997 stars with a total of 1,235 planetary candidates that show transit-like signatures, all with some follow-up work that could not rule out the planet hypothe-
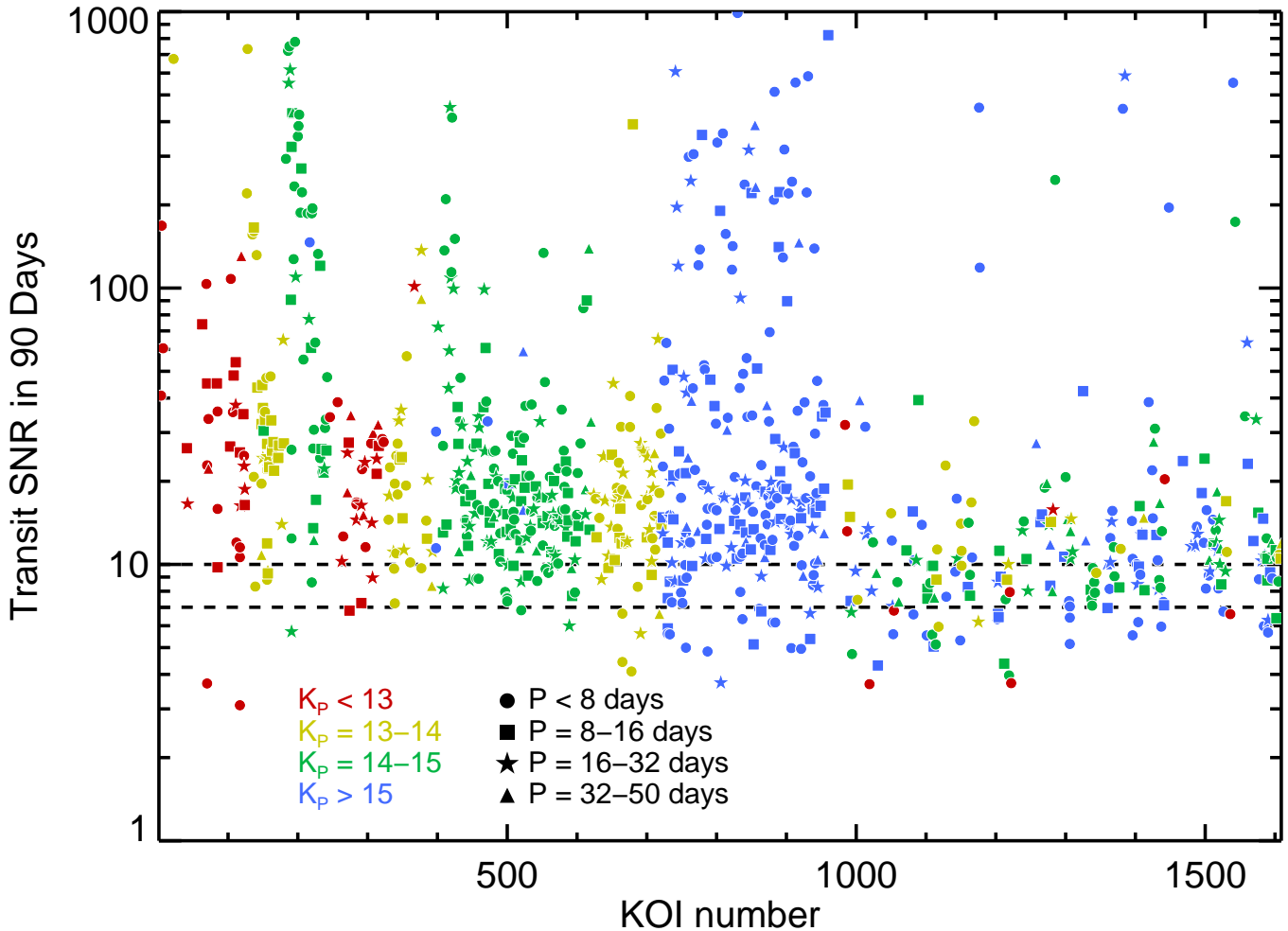
FIG. 2.— Signal to Noise Ratio (SNR) of detected transits during a 90 day quarter. See text for definition of SNR. Only planets orbiting stars with $T_{\rm eff}$ = 4100–6100 K and log g = 4.0–4.9 are shown. Planets orbiting stars with Kepler magnitudes Kp <13, 13–14, 14–15, and > 15 are shown in red, yellow, green, and blue, respectively. Planets with orbital periods P < 8, 8–16, 16–32, and > 32 days are shown as filled circles, squares, five-point stars, and triangles, respectively. Our analysis considers only transits with SNR > 10 (upper dashed line).

sis (Gautier et al. 2010). Borucki et al. (2011) includes three planets discovered in the *Kepler* field before launch: TrES-2b (O'Donovan et al. 2006), HAT-P-7b (Pál et al. 2008), and HAT-P-11b (Bakos et al. 2010a). We are including only those planet candidates that meet two SNR standards: They must have SNR > 10 in one quarter alone and they must have SNR > 7 in all quarters. The former standard should guarantee the latter, but this double-standard reinforces the quality of the planet candidates.

As this data release contains 136 days of photometric data, with only a few small windows of down time, most planet candidates with periods under 50 days have exhibited two or more transits. The multiple transits for P < 50 days offer relatively secure candidates, periods, and radii, provided by the repeated transit light curves. For P < 40 days, *Kepler* has detected typically three or more transits in the publicly available data. Moreover, in Borucki et al. (2011) the periods, radii, and ephemerides are based on the full set of *Kepler* data obtained in Q0–5, constituting over one year of photometric data. Thus, planet candidates with periods under 50 days are securely detected with multiple transits. They have improved SNR in the light curves from the full set of data available

to the *Kepler* team, offering excellent verification, radii, and periods for short period planets.

2.4. *False Positives*

We expect that some of the planet candidates reported in Borucki et al. (2011) are actually false positives. These would be mostly background eclipsing binaries diluted by the foreground star. They may also be background stars orbited by a transiting planet of larger radius, but diluted by the light of the foreground star mimicking a smaller planet. False positives can also occur from gravitationally bound companion stars that are eclipsing binaries or have larger transiting planets. We expect that false positive probabilities will be estimated for most planet candidates in Borucki et al. (2011) using "BLENDER" (Torres et al. 2011).

In the mean time, the false positive rate has been estimated carefully by Morton & Johnson (2011). They find the false positive probability for candidates that pass the standard vetting gates to be less than 10% and normally closer to 5%. In particular, the *Kepler* vetting process included a difference analysis between CCD images taken in and out of transit, allowing direct detection of the pixel that contains the eclipsing binary, if any. This vetting
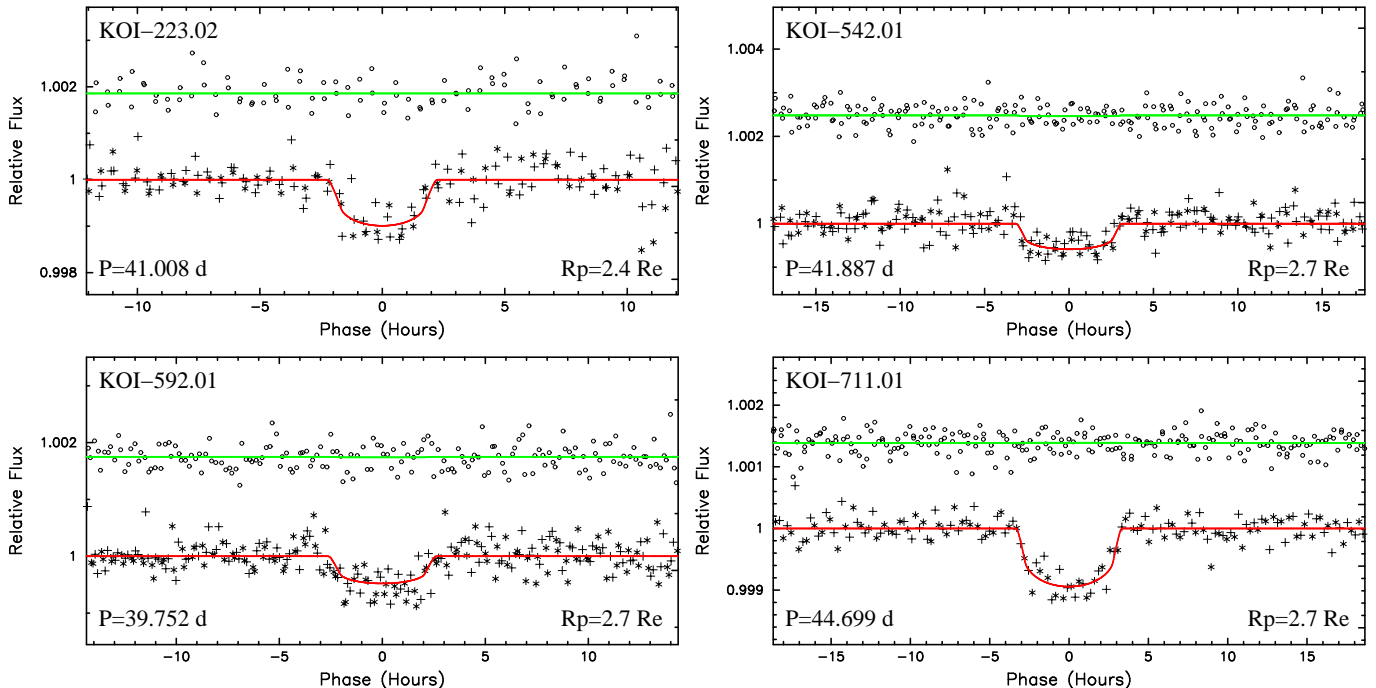
FIG. 3.— Four representative light curves of planet candidates with $R_p = 2$–$3\ R_\oplus$ and $P = 30$–50 days, the domain of most challenging detection in this study. See Table 2 for planetary and stellar properties. In each panel, the transit light curve (lower, red trace) and photometric measurements 180 degrees out of phase (upper, green trace) are shown with the best-fit model overlaid. Plus and star symbols show alternating transits. Only photometry from Q2 is displayed. These light curves demonstrate the data quality for the lowest SNR planet candidates included in this study, most with a transit depth of only ∼10 times greater than the uncertainty in the mean depth due to noise. Still, the transits are clearly visible to the eye. The *Kepler* pipeline is unlikely to miss many of these planet candidates, despite their being in the least detectable domain of the study. This indicates the security of these detections and the high completeness of such planet candidates, in support of Figure 2.

process found that ∼12% of the original planet candidates were indeed eclipsing binaries in neighboring pixels, and these were deemed false positives and removed from Table 2 of Borucki et al. (2011). This process leaves only the one pixel itself, with a half-width of 2 arcsec within which any eclipsing binary must reside. As 12% of the planet candidates had an eclipsing binary within the ∼10 pixels total of the photometric aperture, the rate of eclipsing binaries hidden behind the remaining one pixel is likely to be ∼1.2%, a small probability of false positives. The bound, hierarchical eclipsing binaries were estimated by Morton & Johnson (2011), finding another few percent may be such false positives, yielding a total false positive probability of ∼5–10%. Morton & Johnson (2011) note that the false positive probability depends on transit depth δ, galactic latitude b, and Kp. Using their "detailed framework" and computing the false positive probability for each of the 438 planet candidates among our "solar subset" (Table 1), we estimate that 22 planet candidates are actually false positives.[32] The resulting false positive rate of 5% is on the low end of the 5–10% estimate above because we restricted our stellar sample to bright main sequence stars and planet sample to $R_p > 2\ R_\oplus$. We do not expect this low false positive

rate to substantially impact our statistical results below.

Nearly all of the KOIs reported in Borucki et al. (2011) are formally "planet candidates", absent planet validation (Torres et al. 2011) or mass determination (Borucki et al. 2010b; Koch et al. 2010a; Dunham et al. 2010; Latham et al. 2010; Jenkins et al. 2010a; Holman et al. 2010; Batalha et al. 2011; Lissauer et al. 2011a). For simplicity we will refer to all KOIs as "planets", bearing in mind that a small percentage will turn out to be false positives.

3. PLANET OCCURRENCE

We define planet occurrence, f, as the fraction of a defined population of stars (in $T_{\rm eff}$, $\log g$, Kp) having planets within a domain of planet radius and period, including all orbital inclinations. We computed planet occurrence as a function of planet radius and orbital period in the grid of cells in Figure 4. Within each cell we counted the number of planets detected by *Kepler* for the subset of stars surveyed with sufficient precision to compute the local planet occurrence, $f_{\rm cell}$. Our treatment corrects for planets not detected by *Kepler* because of non-transiting orbital inclinations and because of insufficient photometric precision.

The average planet occurrence within a confined cell of R_p and P is

$$f_{\rm cell} = \sum_{j=1}^{n_{\rm pl,cell}} \frac{1/p_j}{n_{\star,j}}, \qquad (2)$$

where the sum is over all detected planets within the cell that have SNR > 10. In the numerator, $p_j = (R_\star/a)_j$

---

[32] We note that while the precise details of these estimates depend on *a priori* assumptions of the overall planet occurrence rate (which we conservatively take to be 20%) and of the planet radius distribution (which follows Figure 5 of Morton & Johnson (2011)), the overall low false positive probability is controlled by the relative scarcity of blend scenarios compared to planets. We also note that these estimates do not account for uncertainties in $R_\star$, which may result in some jovian-sized candidates actually being M dwarfs eclipsing subgiant stars.
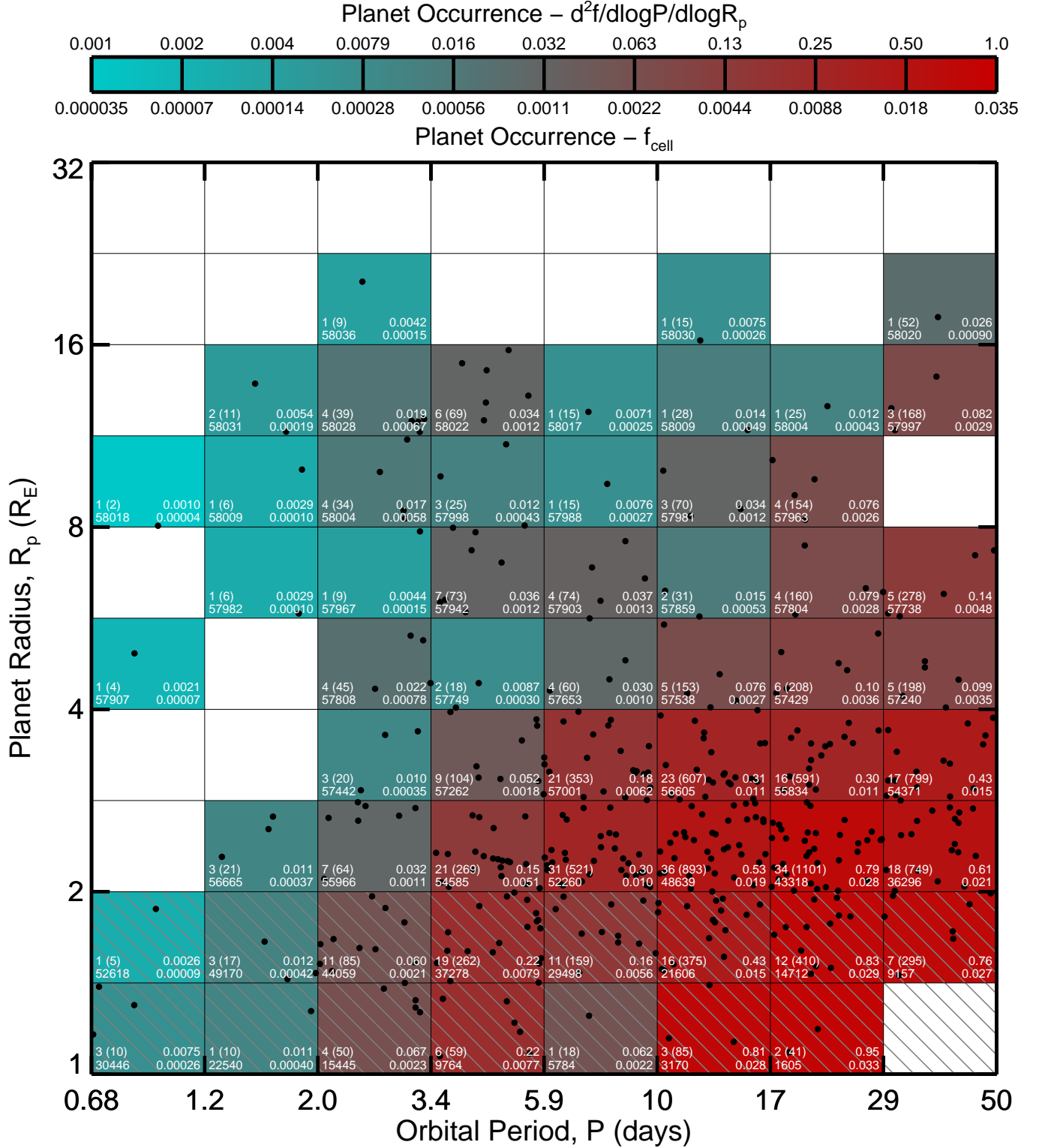
FIG. 4.— Planet occurrence as a function of planet radius and orbital period for $P < 50$ days. Planet occurrence spans more than three orders of magnitude and increases substantially for longer orbital periods and smaller planet radii. Planets detected by *Kepler* having SNR > 10 are shown as black dots. The phase space is divided into a grid of logarithmically spaced cells within which planet occurrence is computed. Only stars in the "solar subset" (see selection criteria in Table 1) were used to compute occurrence. Cell color indicates planet occurrence with the color scale on the top in two sets of units, occurrence per cell and occurrence per logarithmic area unit. White cells contain no detected planets. Planet occurrence measurements are incomplete and likely contain systematic errors in the hatched region ($R_p < 2 R_\oplus$). Annotations in white text within each cell list occurrence statistics: *upper left*—the number of detected planets with SNR > 10, $n_{\rm pl,cell}$, and in parentheses the number of augmented planets correcting for non-transiting geometries, $n_{\rm pl,aug,cell}$; *lower left*—the number of stars surveyed by *Kepler* around which a hypothetical transiting planet with R_p and P values from the middle of the cell could be detected with SNR > 10; *lower right*—$f_{\rm cell}$, planet occurrence, corrected for geometry and detection incompleteness; *upper right*—$d^2 f/d \log_{10} P \, d \log_{10} R_p$, planet occurrence per logarithmic area unit ($d \log_{10} P \, d \log_{10} R_p = 28.5$ grid cells).

is the *a priori* probability of a transiting orientation of planet j. Each individual planet is augmented in its contribution to the planet count by a factor of $a/R_\star$ to account for the number of planets with similar radii and periods that are not detected because of non-transiting geometries. For each planet, its specific value of $(a/R_\star)_j$ is used, not the average $a/R_\star$ of the cell in which it resides. Each scaled semi-major axis $(a/R_\star)_j$ is measured directly from *Kepler* photometry and is not the ratio of two quantities, a_j and $R_{\star,j}$, separately measured with lower precision. In the denominator, $n_{\star,j}$ is the number of stars whose physical properties and photometric stability are sufficient so that a planet of radius $R_{p,j}$ and period P_j would have been detected with SNR > 10 as defined by equation (1). Note that our requirement for SNR > 10 is applied to the numerator (the planets that count toward the occurrence rate) and the denominator (the stars around which those planets could have been detected) of equation (2).

While Figure 4 does not show error estimates for $f_{\rm cell}$, we compute them with binomial statistics and use them in the analysis that follows. We calculate the binomial probability distribution of drawing $n_{\rm pl,cell}$ planets from $n_{\star,{\rm eff,cell}} = n_{\rm pl,cell}/f_{\rm cell}$ "effective" stars. The $\pm 1\sigma$ errors in $f_{\rm cell}$ are computed from the 15.9 and 84.1 percentile levels in the cumulative binomial distribution. Note that $n_{\rm pl,cell}$ is typically a small number (in Figure 4, $n_{\rm pl,cell}$ has a range of 1–36 detected planets) so the errors within individual cells can be significant. These errors and the corresponding occurrence fluctuations between adjacent cells average out when cells are binned together to compute occurrence as a function of radius or period. Also note that our error estimates only account for random errors and not systematic effects.

Figure 4 contains numerical annotations to help digest the wealth of planet occurrence information. In the lower left of each cell is $n_{\star,{\rm mid-cell}}$, the number of *Kepler* targets with sufficient $\sigma_{\rm CDPP}$ such that a central transit of a planet with $R_{\rm p}$ and P values from the middle of the cell could have been detected with SNR > 10. Above this, we list $n_{\rm pl,cell}$ followed by $n_{\rm pl,aug,cell}$ in parentheses. $n_{\rm pl,aug,cell}$ is the total extrapolated number of planets in each cell after correcting for the *a priori* transit probability for each planet,

$$ n_{\rm pl,aug,cell} = \sum_{j=1}^{n_{\rm pl}} 1/p_j. \qquad (3) $$

The annotation in the lower right of each cell is $f_{\rm cell}$. The reader can quickly check that planet occurrence is computed correctly by verifying that $f_{\rm cell} \approx n_{\rm pl,aug,cell}/n_{\star,{\rm mid-cell}}$; planet occurrence is the ratio of the number of planets to the number of stars searched.[33] Finally, the annotation in the top right of each cell is

---

[33] This approximate expression for $f_{\rm cell}$ breaks down in cells where the number of stars with SNR > 10 ($n_\star$) varies rapidly across the cell. Equation (2) computes planet occurrence locally using $n_\star$ for the specific radius and period of each detected planet, while the $n_{\star,{\rm mid-cell}}$ listed in the annotations applies to $R_{\rm p}$ and P in the middle of the cell. Thus, planet occurrence is more poorly determined in regions of Figure 4 where the detection completeness varies rapidly and/or the detected planets are clustered in one section of the cell. These poorly measured regions typically have $R_{\rm p} < 2\ R_\oplus$ and longer orbital periods.

TABLE 3
Planet Multiplicity vs. Planet Size

$R_{\rm p}$ ($R_\oplus$)	Fraction of planet hosts with a second planet …		
	in same $R_{\rm p}$ range	within $\frac{1}{2}R_{\rm p}$–$2R_{\rm p}$	with any $R_{\rm p}$
1.0–1.4	0.05	0.16	0.26
1.4–2.0	0.09	0.25	0.27
2.0–2.8	0.08	0.23	0.25
2.8–4.0	0.12	0.28	0.30
4.0–5.6	0.04	0.09	0.13
5.6–8.0	0.04	0.09	0.13
8.0–11.3	0.00	0.06	0.06
11.3–16.0	0.00	0.00	0.06
16.0–22.6	0.00	0.00	0.00

$f_{\rm cell}$ in units of occurrence per $\mathrm{d}\log_{10} P\ \mathrm{d}\log_{10} R_p$ (occurrence per factor of ten in $R_{\rm p}$ and P), a unit that is independent of the choice of cell size. There are 28.5 grid cells per unit of $\mathrm{d}\log_{10} P\ \mathrm{d}\log_{10} R_p$; that is, a region whose edges span factors of ten in $R_{\rm p}$ and P has 28.5 grid cells of the size shown in Figure 4. Each cell spans a factor of $\sqrt{2}$ in $R_{\rm p}$ and a factor of $5^{1/3}$ in P.

The distribution of planet occurrence in Figure 4 offers remarkable clues about the processes of planet formation, migration, and evolution. *Planet occurrence increases substantially with decreasing planet radius and increasing orbital period.* Planets larger than 1.5 times the size of Jupiter ($R_{\rm p} > 16\ R_\oplus$) are extremely rare. Planets with $P \lesssim 2$ days are similarly rare. Because of incompleteness, we tread with caution for planets with $R_{\rm p} = 1$–$2\ R_\oplus$, but note that these planets have a occurrence similar to planets with $R_{\rm p} = 2$–$4\ R_\oplus$. Their actual occurrence could be higher due to incompleteness of the pipeline at identifying the smallest planets or lower due to a higher rate of false positives.

Planet multiplicity complicates our measurements of planet occurrence. We interpret $f_{\rm cell}$ as the fraction of stars having a planet in the narrow range of P and $R_{\rm p}$ that define a particular cell. With few exceptions, stars are not orbited by planets with nearly the same radii and periods. However, when we apply equation (2) to larger domains of the radius-period plane, for example by marginalizing over P (Section 3.1) or over $R_{\rm p}$ (Section 3.2), the same star can be counted multiple times in equation (2) if multiple planets fall within that larger domain of $R_{\rm p}$ and P. Thus, our occurrence measurements are actually of the mean number of planets per star meeting some criteria, rather than than fraction of stars having at least one planet that meet those criteria. When the rate of planet multiplicity within a domain is low, these two quantities are nearly equal.

The 438 planets in our solar subset of stars (Table 1) orbit a total of 375 stars. The fraction of planets in multi-transiting systems is 0.27 and the fraction of host stars with multiple transiting planets is 0.15. In Table 3 we list three measures of planet multiplicity for the planetary systems within the solar subset (Table 1). For each of the $R_{\rm p}$ ranges in Figure 4 we list the fraction of hosts stars with more than one planet in the specified $R_{\rm p}$ range, the fraction of hosts with one planet in the $R_{\rm p}$ range and a second planet with a radius within a factor of two of the first planet's, and the fraction with one planet in the $R_{\rm p}$ range and a second planet having any $R_{\rm p}$.

Table 3 suggests that multiplicity is common.

Lissauer et al. (2011b) noted that the multi-planet systems observed by Kepler have relatively low mutual inclinations (typically a few degrees) suggesting a significant correlation of inclinations. Converting our measurements of the mean number of planets per star to the fraction of stars having at least one planet requires an understanding of the underlying multiplicity and inclination distributions. Such an analysis is attempted by Lissauer et al. (2011b), but is beyond the scope of this paper.

It is worth identifying additional sources of error and simplifying assumptions in our methods. The largest source of error stems directly from 35% rms uncertainty in $R_\star$ from the KIC, which propagates directly to 35% uncertainty in R_p. We assumed a central transit over the full stellar diameter in equation (2). For randomly distributed transiting orientations, the average duration is reduced to $\pi/4$ times the duration of a central transit. Thus, this correction reduces our SNR in equation (1) by a factor of $\sqrt{\pi/4}$, i.e. a true signal-to-noise ratio threshold of 8.8 instead of 10.0. This is still a very conservative detection threshold. Additionally, our method does not account for the small fraction of transits that are grazing and have reduced significance. We assumed perfect $\sqrt{t}$ scaling for σ_{CDPP} values computed for 3 hr intervals. This may underestimate σ_{CDPP} for a 6 hr interval (approximately the duration of a $P = 50$ day transit) by ~10%. These are minor corrections and affect the numerator and denominator of equation (2) nearly equally.

3.1. *Occurrence as a Function of Planet Radius*

Planet occurrence varies by three orders of magnitude in the radius-period plane (Figure 4). To isolate the dependence on these parameters, we first considered planet occurrence as a function of planet radius, marginalizing over all planets with $P < 50$ days. We computed occurrence using equation (2) for cells with the ranges of radii in Figure 4 but for all periods less than 50 days. This is equivalent to summing the occurrence values in Figure 4 along rows of cells to obtain the occurrence for all planets in a radius interval with $P < 50$ days. The resulting distribution of planet radii (Figure 5) increases substantially with decreasing R_p.

We modeled this distribution of planet occurrence with planet radius as a power law of the form

$$\frac{\mathrm{d}f(R)}{\mathrm{d}\log R} = k_R R^\alpha. \tag{4}$$

Here $\mathrm{d}f(R)/\mathrm{d}\log R$ is the mean number of planets having $P < 50$ days per star in a $\log_{10}$ radius interval centered on R (in $R_\oplus$), k_R is a normalization constant, and α is the power law exponent. To estimate these parameters, we used measurements from the 2–22.7 $R_\oplus$ bins because of incompleteness at smaller radii and a lack of planets at larger radii. We fit equation (4) using a maximum likelihood method (Johnson et al. 2010). Each radius interval contains an estimate of the planet fraction, $F_i = \mathrm{d}f(R_i)/\mathrm{d}\log R$, based on a number of planet detections made from among an effective number of target stars, such that the probability of F_i is given by the binomial distribution

$$p(F_i|n_{\mathrm{pl}}, n_{\mathrm{nd}}) = F_i^{n_{\mathrm{pl}}}(1 - F_i)^{n_{\mathrm{nd}}} \tag{5}$$

where n_{pl} is the number of planets detected in a spec-
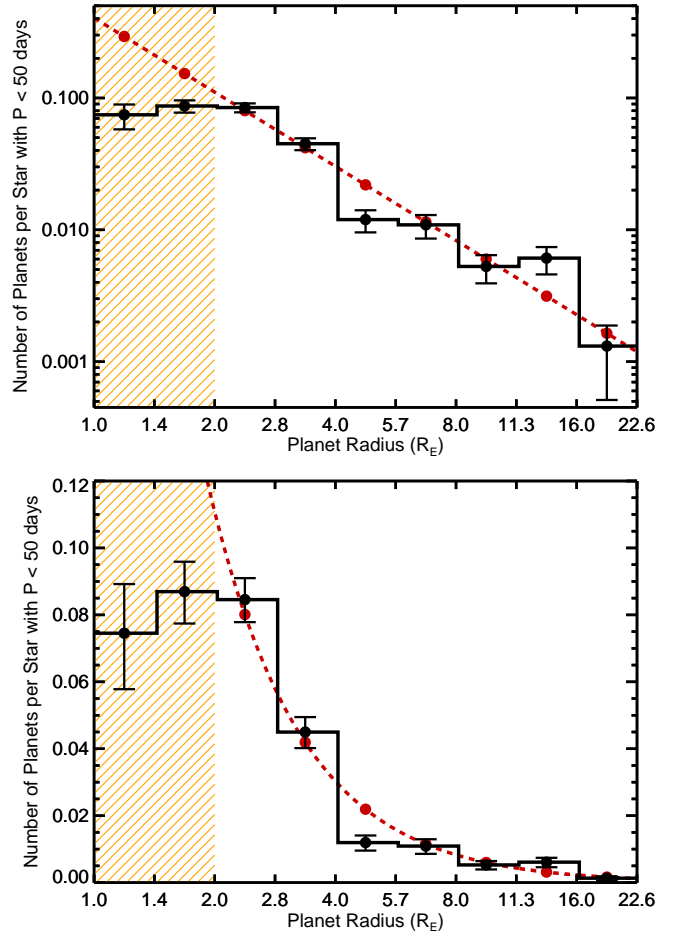

FIG. 5.— Planet occurrence as a function of planet radius for planets with $P < 50$ days (black filled circles and histogram). The top and bottom panels show the same planet occurrence measurements on logarithmic and linear scales. Only GK stars consistent with the selection criteria in Table 1 were used to compute occurrence. These measurements are the sum of occurrence values along rows in Figure 4. Estimates of planet occurrence are incomplete in the hatched region ($R_p < 2\, R_\oplus$). Error bars indicate statistical uncertainties and do not include systematic effects, which are particularly important for $R_p < 2\, R_\oplus$. No planets with radii of 22.6–32 $R_\oplus$ were detected (see top row of cells in Figure 4). A power law fit to occurrence measurements for $R_p = 2$–22.6 $R_\oplus$ (red filled circles and dashed line) demonstrates that close-in planet occurrence increases substantially with decreasing planet radius.

ified radius interval (marginalized over period, $n_{\mathrm{nd}} \equiv n_{\mathrm{pl}}/f_{\mathrm{cell}} - n_{\mathrm{pl}}$ is the effective number of non-detections per radius interval, and f_{cell} is the estimate of planet occurrence over the marginalized radius interval obtained from equation (2). The planet fraction varies as a function of the mean planet radius $R_{p,i}$ in each bin, and the best-fitting parameters can be obtained by maximizing the probability of all bins using the model in equation (4):

$$\mathcal{L} = \prod_{i=1}^{n_{\mathrm{bin}}} p(F(R_{p,i})). \tag{6}$$

In practice the likelihood becomes vanishingly small away from the best-fitting parameters, so we evaluate the logarithm of the likelihood

$$\ln \mathcal{L} = \sum_{i=1}^{n_{\mathrm{bin}}} \ln p(F(R_{p,i})) \tag{7}$$

$$=\sum_{i=1}^{n_{\mathrm{bin}}} n_{\mathrm{pl},i}(\ln k_R + \alpha \ln R_{\mathrm{p},i}) + n_{\mathrm{nd},i}\ln(1 - k_R R_{\mathrm{p},i}^{\alpha}).$$

We calculate $\ln \mathcal{L}$ over a uniform grid in k_R and α. The resulting posterior probability distribution is strongly co-variant in α and k_R. Marginalizing over each parameter, we find $\alpha = -1.92 \pm 0.11$ and $k_R = 2.9^{+0.5}_{-0.4}$, where the best-fit values are the median of the marginalized 1-dimensional parameter distributions and the error bars are the 15.9 and 84.1 percentile levels.

Howard et al. (2010) found a power law planet *mass* function, $\mathrm{d}f/\mathrm{dlog}M = k' M^{\alpha'}$, with $k' = 0.39^{+0.27}_{-0.16}$ and $\alpha' = -0.48^{+0.12}_{-0.14}$ for periods $P < 50$ days and masses $M_{\mathrm{p}}\sin i = 3$–1000 $M_{\oplus}$. We explore planet densities and the mapping of R_{p} to $M_{\mathrm{p}}\sin i$ in Section 5.

3.2. *Occurrence as a Function of Orbital Period*

We computed planet occurrence as a function of orbital period using equation (2). We considered this period dependence for ranges of planet radii ($R_{\mathrm{p}} = 2$–4, 4–8, and 8–32 $R_{\oplus}$). This is equivalent to summing the occurrence values in Figure 4 along two adjacent columns of cells to obtain the occurrence for all planets in specified radius ranges. Figure 6 shows that planet occurrence increases substantially with increasing orbital period, particularly for the smallest planets with $R_{\mathrm{p}} = 2$–4 $R_{\oplus}$.

For $P < 2$ days, planets of all radii in our study (>2 $R_{\oplus}$) are extremely rare with an occurrence of < 0.001 planets per star. Extending to slightly longer orbital periods, hot Jupiters ($P < 10$ days, $R_{\mathrm{p}} = 8$–32 $R_{\oplus}$) are also rare in the *Kepler* survey. We measure an occurrence of only 0.004 ± 0.001 planets per star, as listed in Table 4. That occurrence value is based on Kp < 15 and the other restrictions that define of the "solar subset" (Table 1). Expanding our stellar sample out to Kp < 16, but keeping the other selection criteria constant, we find a hot Jupiter occurrence of 0.005 ± 0.001 planets per star. This fraction is more robust as it is less sensitive to Poisson errors and our concern about detection incompleteness for Kp > 15 vanishes for hot Jupiters that typically produce SNR > 1000 signals. Marcy et al. (2005a) found an occurrence of 0.012 ± 0.001 for hot Jupiters ($a < 0.1$ AU, $P \lesssim 12d$) around FGK dwarfs in the Solar neighborhood (within 50 pc). *Thus, the occurrence of hot Jupiters in the* Kepler *field is only 40% that in the Solar neighborhood.* One might worry that our definition of $R_{\mathrm{p}} > 8$ $R_{\oplus}$ excludes some hot Jupiters detected by RV surveys. For Kp < 16 and the same T_{eff} and $\log g$ criteria, we find an occurrence of 0.0076 ± 0.0013, which is still 40% lower than the RV measurement.

However, we do see modest evidence among the *Kepler* giant planets of the pile-up of hot Jupiters at orbital periods near 3 days (Figures 4 and 6) as is dramatically obvious from Doppler surveys of stars in the Solar neighborhood (Marcy et al. 2008; Wright et al. 2009). These massive, close-in planets are detected with high completeness in both Doppler and *Kepler* techniques (including the geometrical factor for *Kepler*), so the different occurrence values are real. We are unable to explain this difference, although a paucity of metal-rich stars in the *Kepler* sample is one possible explanation. Unfortunately, the metallicities of *Kepler* stars from
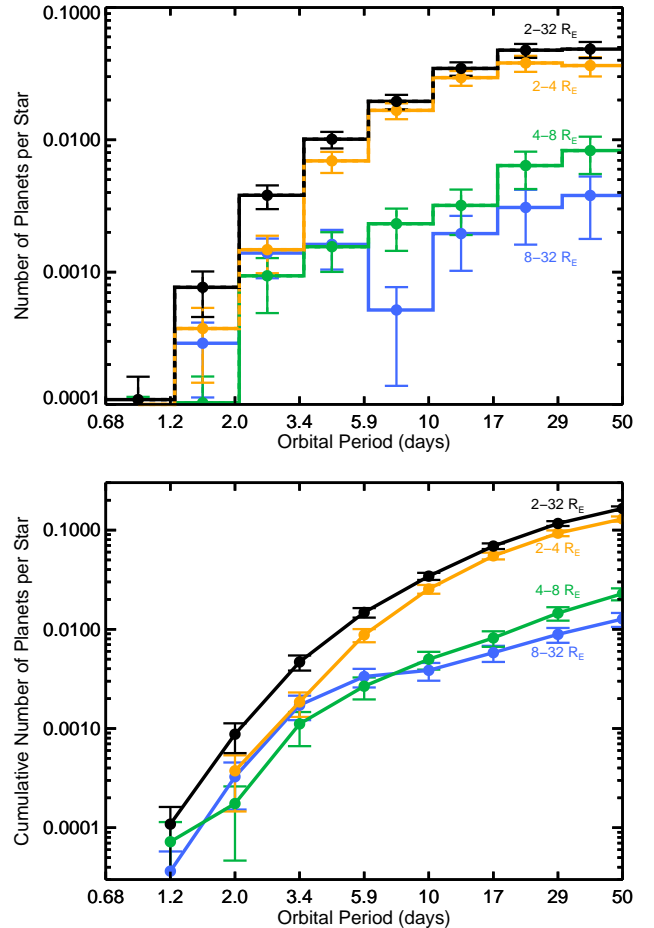

FIG. 6.— Planet occurrence (top panel) and cumulative planet occurrence (bottom panel) as a function of orbital period. The occurrence of planets with radii of 2–32 $R_{\oplus}$ (black), 2–4 $R_{\oplus}$ (orange), 4-8 $R_{\oplus}$ (green), 8–32 $R_{\oplus}$ (blue) are each depicted. Only stars consistent with the selection criteria in Table 1 were used to compute occurrence. Occurrence for planets with $R_{\mathrm{p}} < 2\,R_{\oplus}$ is not shown due to incompleteness. The lower panel (cumulative planet occurrence) is the sum of occurrence values in the top panel out to the specified period.

KIC photometry are inadequate to test this hypothesis (Brown et al. 2011). A future spectroscopic study of *Kepler* stars with LTE analysis similar to Valenti & Fischer (2005) offers a possible test. In additional to the metallicity difference, the stellar populations may have different T_{eff} distributions, despite having similar T_{eff} ranges. Johnson et al. (2010) found that giant planet occurrence correlates with both stellar metallicity and stellar mass (for which T_{eff} is a proxy). A full study of the occurrence of hot Jupiters is beyond the scope of this paper, but we note that other photometric surveys for transiting hot Jupiters orbiting stars outside of the stellar neighborhood have measured reduced planets occurrence (Gilliland et al. 2000; Weldrake et al. 2008; Gould et al. 2006).

The occurrence of smaller planets with radii $R_{\mathrm{p}} = 2$–4 $R_{\oplus}$ rises substantially with increasing P out to ~10 days and then rises slowly or plateaus when viewed in a log-log plot (orange histogram, top panel of Figure 6). Out to 50 days we estimate an occurrence of 0.130 ± 0.008 planets per star. Small planets in this radius range account for approximately three quarters of the planets in our study,

TABLE 4
PLANET OCCURRENCE FOR GK DWARFS

$R_{\rm p}(R_\oplus)$	$P < 10$ days	$P < 50$ days
2–4 $R_\oplus$	0.025 ± 0.003	0.130 ± 0.008
4–8 $R_\oplus$	0.005 ± 0.001	0.023 ± 0.003
8–32 $R_\oplus$	0.004 ± 0.001	0.013 ± 0.002
2–32 $R_\oplus$	0.034 ± 0.003	0.165 ± 0.008

TABLE 5
BEST-FIT PARAMETERS OF CUTOFF POWER LAW MODEL

$R_{\rm p}$ ($R_\oplus$)	k_P	β	P_0 (days)	γ
2–4 $R_\oplus$	0.064 ± 0.040	0.27 ± 0.27	7.0 ± 1.9	2.6 ± 0.3
4–8 $R_\oplus$	0.0020 ± 0.0012	0.79 ± 0.50	2.2 ± 1.0	4.0 ± 1.2
8–32 $R_\oplus$	0.0025 ± 0.0015	0.37 ± 0.35	1.7 ± 0.7	4.1 ± 2.5
2–32 $R_\oplus$	0.035 ± 0.023	0.52 ± 0.25	4.8 ± 1.6	2.4 ± 0.3

corrected for incompleteness.

The occurrence distributions in the top panel of Figure 6 have shapes that are more complicated than simple power laws. Occurrence falls off rapidly at short periods. We fit each of these distributions to a power law with an exponential cutoff,

$$\frac{{\rm d}f(P)}{{\rm d}\log P} = k_P P^\beta \left(1 - e^{-(P/P_0)^\gamma}\right). \qquad (8)$$

This function behaves like a power law with exponent β and normalization k_P for $P \gg P_0$. For periods P (in days) near and below the cutoff period P_0, $f(P)$ falls off exponentially. The sharpness of this transition is governed by γ. Thus the parameters of equation (8) measure the slope of the power law planet occurrence distribution for "longer" orbital periods as well as the transition period and sharpness of that transition.

We fit equation (8) to the four ranges of radii shown in Figure 6 (top panel) and list the best-fit parameters in Table 5. We note that $\beta > 0$ for all planet radii considered, i.e. planet occurrence increases with $\log P$. For the largest planets ($R_{\rm p} = 8$–$32\ R_\oplus$), $\beta = 0.37 \pm 0.35$ is consistent with the power law occurrence distribution derived by Cumming et al. (2008) for gas giant planets with periods of 2–2000 days, ${\rm d}f \propto M^{-0.31\pm0.2} P^{0.26\pm0.1}\, {\rm d}\log M\, {\rm d}\log P$

P_0 and γ can be interpreted as tracers of the migration and stopping mechanisms that deposited planets at the closest orbital distances. With decreasing planet radius, P_0 increases and γ decreases, shifting the cutoff period outward and making the transition less sharp. Thus, gas giant planets ($R_{\rm p} = 8$–$32\ R_\oplus$) on average migrate closer to their host stars (P_0 is small) and the stopping mechanism is abrupt (γ is large). On the other hand, the smallest planets in our study have a distribution of orbital distances (and periods) with a characteristic stopping distance farther out and a less abrupt fall-off close-in.

The normalization constant k_P is highly correlated with the other parameters of equation (8). A more robust normalization is provided by the requirement that
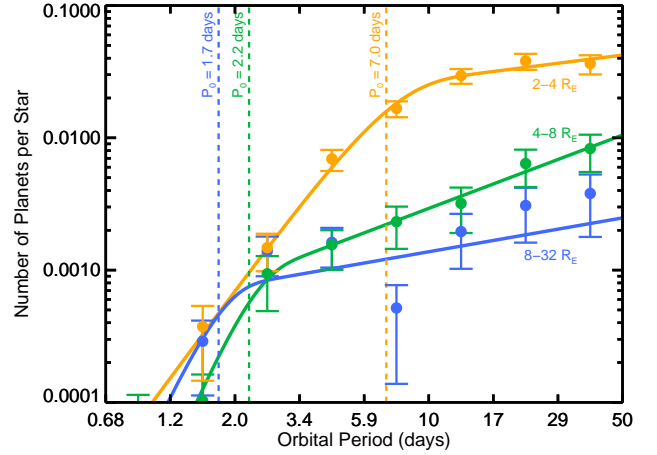

FIG. 7.— Measured planet occurrence (filled circles) as a function of orbital period with best-fit models (solid curves) overlaid. These models are power laws with exponential cutoffs below a characteristic period, P_0 (see text and equation 8). P_0 increases with decreasing planet radius, suggesting that the migration and parking mechanism that deposits planets close-in depends on planet radius. Colors correspond to the same ranges of radii as in Figure 6. The occurrence measurements (filled circles) are the same as in Figure 6, however for clarity the 2–32 $R_\oplus$ measurements and fit are excluded here. As before, only stars in the solar subset (Table 1) and planets with $R_{\rm p} > 2\ R_\oplus$ were used to compute occurrence.

the integrated occurrence to $P = 50$ days is given in Table 4.

4. STELLAR EFFECTIVE TEMPERATURE

4.1. *Planet Occurrence*

In the previous section we considered only GK stars with properties consistent with those listed in Table 1. In particular, only stars with $T_{\rm eff} = 4100$–6100 K were used to compute planet occurrence. Here we expand this range to 3600–7100 K and measure occurrence as a function of $T_{\rm eff}$. This expanded set includes stars as cool as M0 and as hot as F2. For $T_{\rm eff}$ outside of this range there are too few stars to compute occurrence with reasonable errors. We use the same cuts on brightness (Kp < 15) and gravity ($\log g = 4.0$–4.9) as before. We also used the photometric noise $\sigma_{\rm CDPP}$ values (as before) to compute the fraction of target stars around which each detected planet could have been detected with SNR ≥ 10. This ensures that planet detectability down to sizes of 2 $R_\oplus$ will be close to 100%, for all of these included target stars independent of their $T_{\rm eff}$.

We computed planet occurrence using the same techniques as in the previous section, namely equation (2). We subdivided the stars and their associated planets into 500 K bins of $T_{\rm eff}$. We further subdivided the sample by planet radius, considering different ranges of $R_{\rm p}$ (2–4, 4–8, 8–32, and 2–32 $R_\oplus$) separately. In summary, we computed planet occurrence as a function of $T_{\rm eff}$ for several ranges of $R_{\rm p}$ and in all cases we considered all planets with $P < 50$ days.

Figure 8 shows these occurrence measurements as a function of $T_{\rm eff}$. Most strikingly, occurrence is inversely correlated with $T_{\rm eff}$ for small planets with $R_{\rm p} = 2$–$4\ R_\oplus$. Fitting the occurrence of these small planets in the $T_{\rm eff}$ bins shown in Figure 8, we find that a model linear in
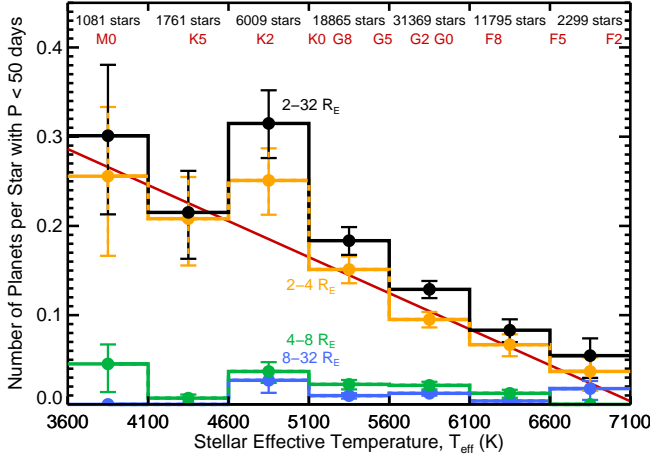
FIG. 8.— Planet occurrence as a function of stellar effective temperature $T_{\rm eff}$. Histogram colors refer to planets with the same ranges of radii as in Figure 6. Here we consider planets with $P < 50$ days and expand beyond the solar subset to $T_{\rm eff} = 3600$–7100 using the same cuts in $\log g$ (4.0–4.9) and Kp (< 15) to select bright main sequence stars. We include only target stars for which photometric noise permits the detection of $2\,R_\oplus$ planets, correcting for reduced detectability of small planets around the larger, hotter stars. The occurrence of small planets ($R_{\rm p} = 2$–$4\,R_\oplus$, orange histogram) rises substantially with decreasing $T_{\rm eff}$. The best-fit linear occurrence model for these small planets is shown as a red line. The number of stars in each temperature bin is listed at the top of the figure. MK spectral types (Cox, A. N. 2000) for main sequence stars are shown for reference.

$T_{\rm eff}$,

$$f(T_{\rm eff}) = f_0 + k_T \left(\frac{T_{\rm eff} - 5100\ {\rm K}}{1000\ {\rm K}} \right), \qquad (9)$$

fits the data well. Using linear least-squares, the best-fit coefficients are $f_0 = 0.165 \pm 0.011$ and $k_T = -0.081 \pm 0.011$ and the relation is valid over $T_{\rm eff} = 3600$–7100 K. We adopted a linear model because it is simple and provides a satisfactory fit with a reduced χ^2 of 1.03. However, we caution that the occurrence measurements in the three coolest bins have relatively large errors and are consistent with a flat occurrence rate, independent of $T_{\rm eff}$.

The occurrence of planets with radii larger than $4\,R_\oplus$ does not appear to correlate with $T_{\rm eff}$ (Figure 8), although detecting such a dependence would be challenging given the lower occurrence of these planets and the associated small number statistics in our restricted sample.

4.2. Sources of Error and Bias

The correlation between the occurrence of 2–4 $R_\oplus$ planets and $T_{\rm eff}$ is striking. In this subsection we consider three possible sources of error and/or bias that could have spuriously produced this result. First, we rule out random errors in the occurrence measurements or in the stellar parameters in the KIC. Next, we consider a systematic bias in $R_\star$, but conclude that any such bias will be too small to cause the correlation. Finally, we consider a systematic metallicity bias as a function of $T_{\rm eff}$. While we consider this unlikely, we cannot rule it out as the cause of the observed correlation.

4.2.1. Random Errors

One might worry that the fit to equation (9) is driven by fluctuations due to small number statistics in the coolest temperature bins. The monotonic trend of rising planet occurrence from 7100 K to 4600 K is less clear for the two coolest bins with $T_{\rm eff} = 3600$–4600 K. The coolest $T_{\rm eff}$ bin, 3600–4100 K, contains only six detected planets and carries the largest uncertainty of any bin. The 4100–4600 K bin contains 13 detected planets. As a test we excluded the hottest and coolest $T_{\rm eff}$ bins and fit equation (9) to the remaining occurrence measurements (4100–6600 K). The best-fit parameters were unchanged to within 1-σ errors.

Next, we checked to see if random or systematic errors in stellar parameters could cause the correlation of 2–4 $R_\oplus$ planet occurrence with $T_{\rm eff}$. The key stellar parameters from the KIC are $T_{\rm eff}$ and $\log g$ which have RMS errors of 135 K and 0.25 dex, respectively. Stellar radii carry fractional errors of 35% RMS stemming from the $\log g$ uncertainties.

Using a Monte Carlo simulation, we assessed the impact of these random errors in the KIC parameters on the noted correlation. In 100 numerical realizations, we added gaussian random deviates to the measured $T_{\rm eff}$ and $\log g$ values for every star in the KIC. These random deviates, $\Delta \log g$ and $\Delta T_{\rm eff}$, had RMS values equal to the RMS errors of their associated variables (135 K and 0.25 dex). Using the new $\log g$ values we updated $R_\star$ for every star using $R_{\star,{\rm new}} = R_{\star,{\rm old}} 10^{\Delta \log g / 2}$. Planet radii, $R_{\rm p}$, were updated in proportion to the change in $R_\star$ for their host stars. With each simulated KIC we performed the entire analysis of this section: we selected KIC stars that meet the $T_{\rm eff}$, $\log g$, and Kp criteria, divided those stars into 500 K subgroups, computed the occurrence of $R_{\rm p} = 2$–$4\,R_\oplus$ planets in each $T_{\rm eff}$ bin using the perturbed $R_{\rm p}$ values, and fit a linear function to the occurrence measurements in each $T_{\rm eff}$ bin yielding f_0 and k_T. The standard deviations of the distributions of f_0 and k_T from the Monte Carlo runs are 0.011 and 0.009, respectively. These uncertainties are nearly equal to the statistical uncertainties of f_0 and k_T quoted above that are derived from the binomial uncertainty of the number of detected planets within each $T_{\rm eff}$ bin. Thus, our quoted errors on f_0 and k_T above probably underestimate the true errors by $\sim \sqrt{2}$. We conclude that the correlation between $T_{\rm eff}$ and the occurrence of 2–4 $R_\oplus$ planets is not an artifact of random errors in KIC parameters.

4.2.2. Systematic $R_\star$ Bias?

Potential systematic errors in the KIC parameters present a greater challenge than random errors. We assessed the impact of systematic errors by considering the null hypothesis—that the occurrence of 2–4 $R_\oplus$ planets is actually independent of $T_{\rm eff}$—and determined how large the systematic error in $R_\star(T_{\rm eff})$ would have to be to produce the observed correlation of occurrence with $T_{\rm eff}$ (equation 9). That is, systematic errors have to account for the factor of 7 increase in the occurrence of 2–4 $R_\oplus$ planets between the $T_{\rm eff} = 6600$–7100 K and 3100–3600 K bins. In this imagined scenario, the photometric determination of $\log g$ in the KIC has a systematic error that is a function of $T_{\rm eff}$. This systematic error causes corresponding errors in $R_\star$ and ultimately $R_{\rm p}$ that depend on $T_{\rm eff}$. We assumed that the power law radius

distribution measured in Section 3.1 is independent of $T_{\rm eff}$ and that it remains valid for $R_{\rm p} < 2\ R_\oplus$. Then the systematic error in $R_{\rm p}$ would shift the bounds of planet radius in each $T_{\rm eff}$ bin. That is, in the lowest $T_{\rm eff}$ bin (3100–3600 K), while we intended to measure occurrence for planets with radii 2–4 $R_\oplus$, we actually measured occurrence over a range of smaller radii, (2–4 $R_\oplus$)/S, where the occurrence rate is intrinsically higher. Here, S is a dimensionless scaling factor that describes the size of the systematic $R_{\rm p}$ error in the $T_{\rm eff} = 3100$–3600 K bin. Similarly, for the $T_{\rm eff} = 6600$–7100 K bin we intend to measure the occurrence of 2–4 $R_\oplus$ planets, but instead we measure the occurrence of planets with $R_{\rm p} = S\cdot(2$–4 $R_\oplus)$ because of systematic errors in $R_{\rm p}(T_{\rm eff}) \propto R_\star(T_{\rm eff})$. Using the power law dependence for occurrence with $R_{\rm p}$ (equation 4), we find that $S = (1/7)^{\alpha/2} = 6.2$ for the systematic error in $R_{\rm p}(T_{\rm eff})$ to cause a factor of 7 occurrence error between the coolest and hottest $T_{\rm eff}$ bins. A factor of 6.2 error in $R_\star$ corresponds to a $\log g$ error of 1.6 dex and is akin to mistaking a subgiant for a dwarf. Surely systematic errors in $R_\star$ and $\log g$ from the KIC are smaller than this. The KIC was constructed almost entirely for the purpose of selecting targets for the planet search by excluding evolved stars. Brown et al. (2011) compared the $\log g$ values from the KIC and LTE spectral synthesis of Keck-HIRES spectra and found that only one star out of 34 tested had a $\log g$ discrepancy of greater than 0.3 dex (see their Figure 8). We reject the null hypothesis and conclude that the strong correlation between the occurrence of 2–4 $R_\oplus$ planets and $T_{\rm eff}$ is real.

4.2.3. *Systematic Metallicity Bias?*

Another potential bias stems from the metallicity gradient as a function of height above the galactic plane (Bensby et al. 2007; Neves et al. 2009). The *Kepler* field sits just above the galactic plane, with a galactic latitude range $b = 6$–20 degrees. The most luminous and hottest stars observed by the magnitude-limited *Kepler* survey are on average the most distant. Because of the slant observing geometry, these stars also have the greatest height above the galactic plane. Likewise, the least luminous and coolest stars observed by Kepler are closer to Earth and only a small distance above the plane. Given that the average metallicity declines with distance from the galactic plane, one might expect that the hottest stars have lower metallicity, on average, than the coolest stars observed by *Kepler*.

This hypothesis suggests a key test: does the occurrence of 2–4 $R_\oplus$ planets depend on [Fe/H]? Unfortunately we are not able to perform this test using stellar parameters from the KIC. While $T_{\rm eff}$ values are accurate to 135 K (rms), [Fe/H] values are of poor quality. Brown et al. (2011) found [Fe/H] errors of 0.2 dex (rms), and possibly higher due to systematic effects. Thus, the [Fe/H] values from the KIC are not helpful in testing the hypothesis that the occurrence of 2–4 $R_\oplus$ planets depends on metallicity.

To get a sense of the size of the metallicity gradient as a function of $T_{\rm eff}$, we simulated our magnitude-limited observations of the *Kepler* field using the Besancon model of the galaxy (Robin et al. 2003). This simulation produced a synthetic set of stars (with individual values of $T_{\rm eff}$, $\log g$, [Fe/H], $M_\star$, etc.) based on the coordinates of the *Kepler* field. We computed the median [Fe/H] for the seven $T_{\rm eff}$ bins in Figure 8 and found, from coolest to hottest, [Fe/H] (median) $= -0.02, -0.03, -0.03, -0.06,$ $-0.07, +0.01, +0.04$. The somewhat surprising upturn in metallicity in the two hottest $T_{\rm eff}$ bins appears to be due to an age dependence with $T_{\rm eff}$; younger stars are more metal rich. The two hottest bins have a median age of 2 Gyr, while the five cooler $T_{\rm eff}$ bins have median ages of 4–5 Gyr. We conclude based on this synthetic galactic model that [Fe/H] varies by perhaps $\sim$0.1 dex over our $T_{\rm eff}$ range and that the dependence need not be monotonic due to age effects.

It is also worth considering how large of an [Fe/H] gradient is needed to increase *giant* planet occurrence by a factor of seven. Clearly, occurrence trends for jovian planets and 2–4 $R_\oplus$ planets need not be similar, but these larger planets offer a sense of scale than may be relevant for smaller planets. For giant planets, Fischer & Valenti (2005) found that occurrence scales as $\propto 10^{2.0[{\rm Fe/H}]}$, while Johnson et al. (2010) found $\propto 10^{1.2[{\rm Fe/H}]}$, after accounting for the occurrence dependence on $M_\star$. These scaling relations suggest that [Fe/H] gradients of 0.4–0.7 dex are needed to affect a factor of seven change in occurrence. A metallicity change of only $\sim$0.1 dex among 2–4 $R_\oplus$ planet hosts seems unlikely to change planet occurrence by the amount we observed. Further, if the occurrence of such planets depends so sensitively on [Fe/H], it seems likely that Doppler surveys of them would have detected this trend among the $\sim$30 RV-detected planets with $M_{\rm p} \sin i < M_{\rm Neptune}$.

The possibility that increased metallicity correlates with increased 2–4 $R_\oplus$ planet occurrence contradicts tentative trends of low-mass planets observed by Doppler surveys. Valenti (2010) noted that among the host stars of Doppler-detected planets, those stars with *only* planets less massive than Neptune are metal poor relative to the Sun. This tentative threshold is intriguing, but it only shows that the distribution of *detected* planets has an apparent [Fe/H] threshold, not that the occurrence of these planets depends systematically on [Fe/H]. To interpret the threshold physically, one needs to check for metallicity bias in the population of Doppler target stars.

5. PLANET DENSITY

It is tempting to extract constraints on the densities of small planets by comparing the distribution of radii measured by *Kepler* to the distribution of minimum masses ($M_{\rm p} \sin i$) measured by Doppler-detected planets from surveys of the Solar neighborhood (Cumming et al. 2008; Howard et al. 2010). This effort may be partially compromised by the different populations of targets stars, despite our efforts to select stars with similar $\log g$ and $T_{\rm eff}$ distributions. The *Kepler* target stars are typically $\sim$50–200 pc above the Galactic plane while Doppler target stars reside typically within 50 pc of the Sun near the plane. Indeed in Section 3.2 we saw that the hot Jupiter occurrence was 2.5 times lower in the *Kepler* survey than in the Doppler surveys, suggesting a difference in stellar populations, possibly related to the decline in metallicity with Galactic latitude and/or differing $T_{\rm eff}$ distributions. Nonetheless, one should not ignore the opportunity to search for information from combining the *Kepler* and Doppler planet occurrences, with caveats prominently in mind.

We first consider known individual planets that

have measured masses, radii, and implied bulk densities. Placing these well-measured planets on theoretical mass-radius relationships (e.g., Valencia et al. 2006; Seager et al. 2007; Sotin et al. 2007; Baraffe et al. 2008; Grasset et al. 2009) provides insight into the range of compositions encompassed by the detected planets. Our goal is to complement these few well-studied cases with statistical constraints on the planet density distribution.

5.1. *Known Planets*

We begin by considering the known planets with $R_p <$ 8 $R_\oplus$ and $M_p > 0.1$ $M_\oplus$. This range of parameters selects planets smaller than Saturn and as large or larger than Mars. Figure 9 shows all such planets with good mass and radius measurements from our solar system and other systems. Theoretical calculations of Kepler-10b (Batalha et al. 2011) based on its mass and radius (4.5 $M_\oplus$ and 1.4 $R_\oplus$) suggest a rock/iron composition with little or no water. Corot-7b has a radius of 1.7 $R_\oplus$ (Léger et al. 2009). Queloz et al. (2009) measured a mass of 4.8 $M_\oplus$ for this planet, implying a density of 5.6 g cm^{-3} and a rocky composition. However, the mass and density have remained controversial. Independent mass determinations based on the same spot-contaminated Doppler data yield masses that vary by a factor of 2–3 (Pont et al. 2011; Hatzes et al. 2010; Ferraz-Mello et al. 2010). We adopt the mass estimate of 1–4 $M_\oplus$ from Pont et al. (2011), which implies a wide range of possible compositions and also marginally favors a water/ice-dominated planet. GJ 1214b is a less dense super-Earth orbiting an M dwarf. The planet has been modeled as a solid core surrounded by H/He/H$_2$O and may be intermediate in composition between ice giants like Uranus and Neptune and a 50% water planet (Nettelmann et al. 2010). The discovery of the six co-planar planets orbiting Kepler-11 added five planets with measured masses (from transit-timing variations) to Figure 9 (Lissauer et al. 2011a). The remaining exoplanets in Figure 9 all have masses greater than Neptune's (17 $M_\oplus$) and densities less than 2 g cm^{-3}: Kepler-4b (Borucki et al. 2010b), Gl 436b (Maness et al. 2007; Gillon et al. 2007; Torres et al. 2008), HAT-P-11b (Bakos et al. 2010a), HAT-P-26b (Hartman et al. 2011), Corot-8b (Bordé et al. 2010), HD 149026b (Sato et al. 2005; Torres et al. 2008).

Figure 9 shows that among known planets their radii increase with planet mass faster than do any of the theoretical curves representing solid compositions of iron, rock, or ice. This rapid increase in radius with mass suggests that planets of higher mass contain larger fractional amounts of H/He gas. The slope increases markedly for masses above 4.5 $M_\oplus$, indicating that above that planet mass the contribution of gas is common, even for these close-in planets. Apparently planets above 4.5 $M_\oplus$ are rarely solid. We suspect that for planets orbiting beyond 0.1 AU where collisional stripping of the outer envelope is less energetic and common, the occurrence of gaseous components will be greater.

Fortney et al. (2007b) modeled solid exoplanets composed of pure water ("ice"), rock (Mg$_2$SiO$_4$), iron, and binary admixtures. Their models include no gas component and are shown as gray lines in Figure 9. Adding gas to any of the models increases R_p and decreases ρ (Adams et al. 2008). Thus planets below and to the right
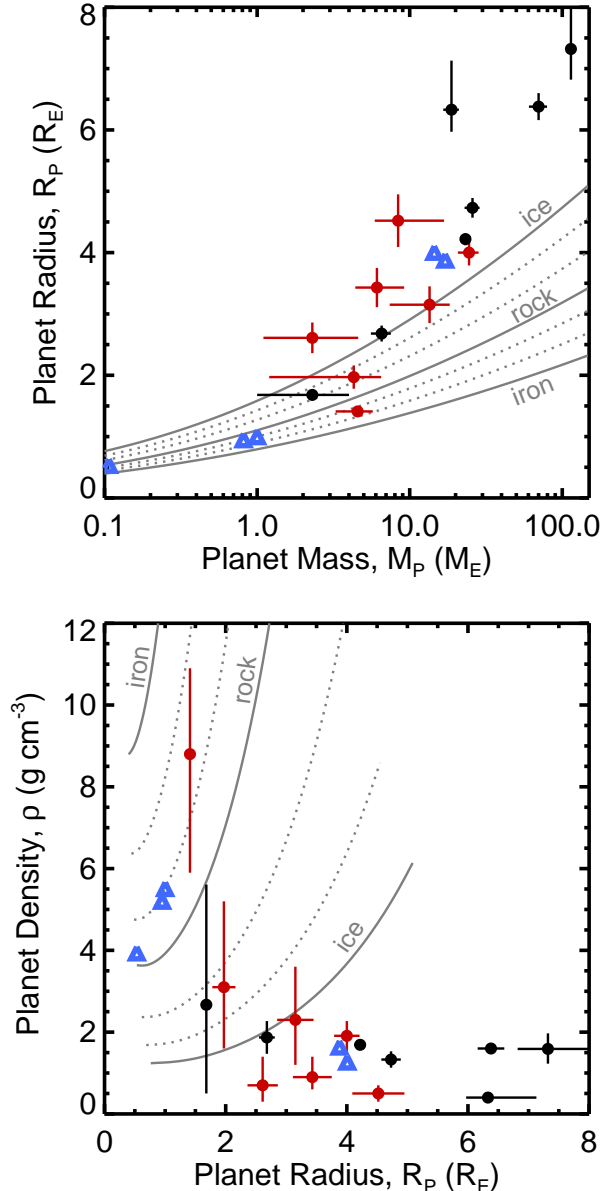

FIG. 9.— Planet masses, radii, and densities for $R_p < 8$ $R_\oplus$ and $M_p > 0.1$ $M_\oplus$. The upper panel shows radius as a function of mass and the lower panel shows density as a function of radius. Solar system planets (Mars, Venus, Earth, Uranus, and Neptune) are depicted as blue triangles. Extrasolar planets (filled circles) are colored red for *Kepler* discoveries and black for discoveries by other programs. Solid gray lines indicate the densities of solid planets composed of pure ice, pure rock, and pure iron using the Fortney et al. (2007a,b) models. Dotted gray lines show the densities of admixture compositions (from bottom to top in lower panel): 67/33% ice/rock, 33/67% ice/rock, 67/33% rock/iron, 33/67% rock/iron.

of the ice contour (Figure 9, lower panel) have low densities due to a gas component. Planets above the ice contour contain increasing fractions of rock and iron, depending on the specific system. Compositional details matter greatly for specific systems, but for our simple purpose we make the crude approximation that planets with $R_p \lesssim 3$ $R_\oplus$ that have $\rho \gtrsim 4$ g cm^{-3} are composed substantially of refractory materials (usually rock in the form of silicates and iron/nickel). These planets may

have some water and gas, but those components do not dominate the planet's composition as they do for Uranus, Neptune, and larger planets.

5.2. *Mapping Kepler Radii to Masses*

The Eta-Earth Survey measured planet occurrence as a function of $M_p \sin i$ in a volume-limited sample of 166 G and K dwarfs using Doppler measurements from Keck-HIRES. The stars have a nearly unbiased metallicity distribution and are chromospherically quiet to enable high Doppler precision. In all, 35 planets were detected around 24 of the 166 stars, including super-Earths and Neptune-mass planets (Howard et al. 2009, 2011a,b). Correcting for inhomogeneous sensitivity at the lowest planet masses, Howard et al. (2010) measured increasing planet occurrence with decreasing mass over five planet mass domains, $M_p \sin i$ = 3–10, 10–30, 30–100, 100–300, 300–1000 $M_\oplus$, spanning super-Earths to Jupiter-mass planets. This study was restricted to planets with $P < 50$ days.

We mapped the planet radius distribution from *Kepler* (Figure 4, including planets down to 1 $R_\oplus$) onto mass ($M_p \sin i$) using toy density functions, $\rho(R_p)$. These single-valued functions map all planets of a particular radius, R_p, onto a planet mass $M_p = 4\pi\rho(R_p)R_p^3/3$. Of course, real planets exhibit far more diversity in radii for a given mass owing to different admixtures of primarily iron/nickel, rock, water, and gas. Nevertheless, the models allow us to check if average masses associated with *Kepler* radii are consistent with Doppler measurements.

As part of this numerical experiment we converted M_p to $M_p \sin i$ for each simulated planet using random orbital orientations (inclinations i drawn randomly from a probability distribution function proportional to $\sin i$.) Our simulated $M_p \sin i$ distributions account for the transit probabilities of planets detected by *Kepler* and the detection incompleteness for planets with small radii. That is, the simulated $M_p \sin i$ distributions reflect the true distribution of planet radii (Section 3.1).

Figure 10 shows simulated $M_p \sin i$ distributions assuming several toy density functions. These distributions are binned in the same $M_p \sin i$ intervals as in the Howard et al. (2010) study. In the left column $\rho(R_p) = \rho_0$, where ρ_0 is a constant. From bottom to top, we considered four densities, ρ_0 = 0.4, 1.35, 1.63, and 5.5 g cm^{-3} (the bulk densities of HAT-P-26b, Jupiter, Neptune, and Earth). We are most interested in the densities of small planets so we make comparisons in the two lowest mass bins for which Eta-Earth Survey measurements are available, $M_p \sin i$ = 3–10 and 10–30 $M_\oplus$. In these bins, the predicted occurrence from *Kepler* is too small by 1.5–2σ compared with the Eta-Earth Survey measurements for the three lowest constant density models, ρ_0 = 0.4, 1.35, and 1.63 g cm^{-3}. *Kepler* predicts fewer small planets than the Eta-Earth Survey measured. The simulated $M_p \sin i$ distribution matches the observed $M_p \sin i$ distribution well for an assumed density, $\rho = 5.5$ g cm^{-3}. While this model is clearly unphysical when extended over the entire radius range, consistency in the two lowmass bins suggests that the small planets have higher densities.

We explored slightly more complicated density functions in the right column of Figure 10. These functions are piece-wise constant density models, with density rising to 4.0, 5.5, and 8.8 g cm^{-3} for small radii, as depicted in the sub-panels of Figure 10. (Kepler-10b has a density of 8.8 g cm^{-3}; Batalha et al. 2011.) We find the greatest consistency between the synthetic and measured mass distributions for two density models. One (model h) is shown in the upper right panel of Figure 10 which has ρ = 8.8, 5.5, 1.64, 1.33 g cm^{-3} for R_p = 1–1.4, 1.4–3.0, 3.0–6.0, and >6.0 $R_\oplus$, respectively. This model has a high density (8.8 g cm^{-3}) for the smallest planets but successively smaller densities for larger planets, approximately consistent with the densities of known planets in Figure 9. The other successful model (g) has a density of 4 g cm^{-3} for the smallest planets, with declining densities for larger planets, qualitatively similar to the previous model (h). This model (g) also yields a predicted distribution of $M_p \sin i$ that agrees well with the observed distribution of $M_p \sin i$. Thus it too is viable. Both successful models, g and h, are characterized by a high density for the smallest planets of 1–3 $R_\oplus$. We tried a variety of piecewise constant density functions and found that all models that achieved consistency ($< 1\sigma$ difference in the 3–10 and 10–30 $M_\oplus$ bins) have $\rho \gtrsim 4$ g cm^{-3} for $R_p \lesssim 3$ $R_\oplus$.

5.3. *Conclusions*

The mapping of radius to mass offers circumstantial evidence that a substantial population of small planets detected by *Kepler* have high densities. Rocky composition for the smallest planets supports the core-accretion model of planet formation (Pollack et al. 1996; Lissauer et al. 2009; Movshovitz et al. 2010). But we caution again that the stellar populations of the *Kepler* and Doppler surveys may be quite different. Planet multiplicity also makes this an especially challenging comparison. We computed the simulated *Kepler* mass distributions (black histograms in Figure 10) based on occurrence measured as the average number of planets per star while the Doppler results from the Eta-Earth Survey (red points in Figure 10) computed occurrence as the fraction of stars hosting at least one planet in the specified $M_p \sin i$ interval. This difference is based on intrinsic limitations of each approach. To infer the fraction stars with at least one planet from a transit survey requires an assumption about the mutual inclinations (Lissauer et al. 2011b). For Doppler surveys, it is significantly easier to determine if a particular star has at least one planet down to some specified mass limit, but it is much more difficult to be sure that all planets orbiting a star have been detected down to that same mass limit (Howard et al. 2010). Finally, we note that no planets at the extreme of our proposed high density regime ($R_p \sim 3$ $R_\oplus$ and $\rho \sim 4$ g cm^{-3}) have been detected (Figure 10). To date all detected planets with $R_p > 2$ $R_\oplus$ have $\rho < 2$ g cm^{-3}. We conclude that while this technique offers qualitative support for rising density with decreasing planet size, in practice extracting firm quantitative conclusions is difficult because of the intrinsic differences between Doppler and transit searches.
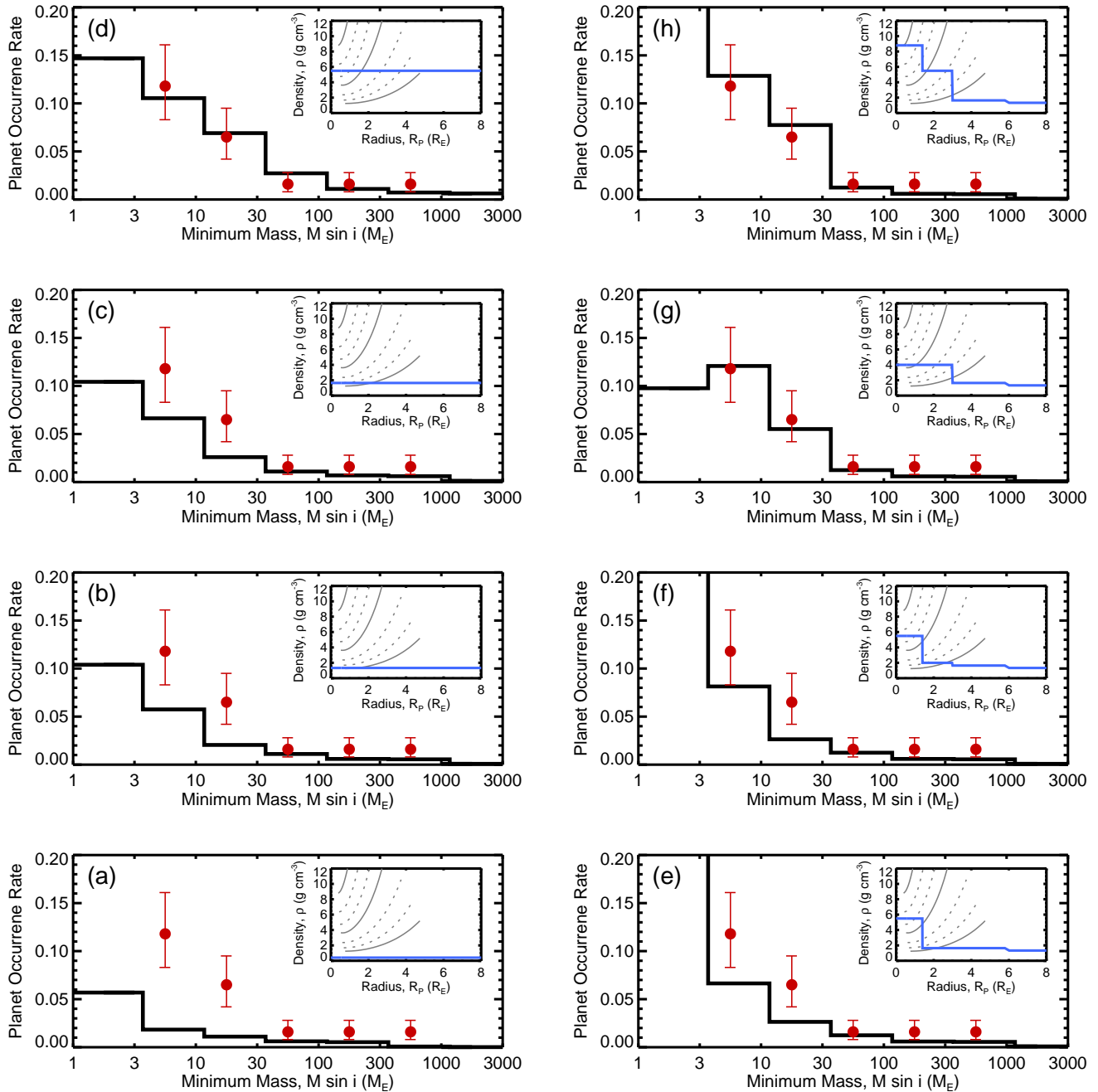
6. DISCUSSION

6.1. *Methods*

Fɪɢ. 10.— Predicted mass distributions ($M_P \sin i$, black histograms) based on planet radii measured by *Kepler* and hypothetical planet density functions (blue lines in inset panels). The Fortney et al. (2007a,b) theoretical density curves for solid planets from Figure 9 are plotted as solid and dotted gray lines in each inset panel. Planet occurrence measurements as a function of $M_P \sin i$ from the Eta-Earth Survey (Howard et al. 2010) are shown as red filled circles. Panels in the left column show the mass distributions resulting from toy constant-density models. From bottom to top (panels a–d), all planets have densities of 0.4, 1.35, 1.63, and 5.5 g cm^{-3}, independent of radius, in analogy with the densities of Earth, Neptune, Jupiter, and HAT-P-26b. In the right column (panels e–h) density increases with decreasing planet radius, as depicted by the inset density functions. Density functions that increase above ~4.0 g cm^{-3} for planets with $R_P \lesssim 3 \, R_\oplus$ yield greater consistency between the Eta-Earth Survey and *Kepler*.

We have attempted to measure pristine properties of planets that can be compared with, and can inform, theories of the formation, dynamical evolution, and interior structures of planets. We have built upon the unprecedented compendium of over 1200 planet candidates found by the historic *Kepler* mission (Borucki et al. 2011). One goal here was to measure planet occurrence—the number of planets per star having particular orbital periods and planet radii—by minimizing the deleterious effects of detection efficiencies that are a function of planet properties, notably radius and orbital period.

Our treatment of the vast numbers of target stars and transiting planet candidates involved careful accounting of two important effects. First, only planets whose orbital planes are nearly aligned to *Kepler*'s line of sight will transit their host star, leaving many planets undetected. We applied the standard geometrical correction for the small probability, $R_\star/a$ in equation (2), that the orbital plane is sufficiently aligned to cause a transit. In counting planets, we assumed that for each detected planet candidate there are actually $a/R_\star$ planets, on average, at all inclinations. Second, only planets whose transits produce photometric signals exceeding some signal-to-noise threshold will be reliably detected. For each possible planet radius and orbital period, we carefully identified the subset of the *Kepler* target stars *a priori* around which such planets could be detected with high probability. We adopted a threshold SNR of 10 for the transit signal in a single 90 day quarter of data, thereby limiting both the target stars and the planet detections with this SNR threshold. To be included, a target star must have a radius and photometric noise that allowed a planet detection with SNR > 10, i.e. a transit depth 10 times greater than the uncertainty in the mean depth from noise. Such restricted target stars offer a high probability that planets will be detected.

We further selected *Kepler* target stars having a specific range of $T_{\rm eff}$, $\log g$, and brightness to ensure a well defined sample of stars. We consider only bright target stars (Kp < 15). We ignore all other Kepler target stars and their associated planets. Remarkably, this *a priori* selection of *Kepler* target stars immediately yields a sample of only ∼58,000 stars (and fewer when accounting for requisite photometric noise), not the full 156,000 stars. For most of the paper, we restricted the sample to main sequence G and K stars ($\log g$ = 4.0–4.9, $T_{\rm eff}$ = 4100–6100 K) to permit comparison with similar Sun-like stars in the Eta-Earth Survey. This selection of *Kepler* target stars for a given planet radius and orbital period crucially leaves only a subset of stars in the "sub-survey" for those planet properties. Importantly, for planets with small radii (near 2 $R_\oplus$) and long periods (near 50 days), only some 36,000–49,000 stars are amenable to detection of such difficult-to-detect planets, as shown in the the annotations in the lower left corners of the cells in Figure 4. By counting planets and dividing by the number of appropriate stars that *could* have permitted their secure detection we computed the planets per star for a specific planet radius and orbital period (within a specified delta in each quantity).

6.2. *Comparison with Borucki et al. (2011)*

It is worth describing the differences between this paper and Borucki et al. (2011) resulting from differing goals and methods. The primary propose of Borucki et al. (2011) was to summarize the results of the *Kepler* observations and to act as a guide to the tables of data. In doing so they tripled the number of known planets (even when allowing for a false positive rate of ∼5%; Morton & Johnson 2011). Borucki et al. (2011) considered the *number distributions* of all planets detected by *Kepler*, independent of the properties of their host stars ($T_{\rm eff}$, $\log g$, Kp, $\sigma_{\rm CDPP}$). They also computed the "intrinsic frequencies" of planetary candidates, a close cousin of our planet occurrence measurements, and plotted these frequencies as a function of $T_{\rm eff}$.

The results in this paper are derived directly from the planets announced in Borucki et al. (2011) and from stellar parameters in the KIC (Brown et al. 2011). We measure the occurrence distributions of planets orbiting bright, main sequence G and K stars, which represent only a third of the stars observed by *Kepler* and considered in Borucki et al. (2011). Our desire for high detection completeness compelled us to consider only robustly detected planets satisfying $R_{\rm p} > 2\ R_\oplus$, $P < 50$ days, SNR > 10 in 90 days of photometry, and stars with Kp < 15. This selection of stars and planets facilitated comparison with the Eta-Earth Survey (Howard et al. 2010), which focused on the Doppler detection of planets orbiting G and K dwarfs with $P < 50$ days. In this paper we measured the detailed patterns of planet occurrence as a function of $R_{\rm p}$, P, and $T_{\rm eff}$, only for that subset of stars and interpreted these distributions in the context of planet formation, evolution, and composition.

Borucki et al. (2011) chose to compute intrinsic frequencies in small domains of semi-major axis and planet radius while we work in a space of orbital periods and planet radii. There are trade-offs with these choices. We chose to work in period space because *Kepler* directly measures orbital periods and translating to semi-major axes requires either assumed stellar masses or radii. On the other hand, by working in small domains of semi-major axis, Borucki et al. (2011) compensate for this by considering the range of orbital periods and transit durations that contribute to each domain for the range of masses and radii among the target stars. In this paper we applied a binary detection criterion of SNR > 10 for 90 days of photometry (approximately one quarter). Borucki et al. (2011) adopted a detection criterion of SNR > 7 for the 136 days of Q0–Q2, with corrections for the probability of low SNR detections (e.g. 7-σ detections are only recognized 50% of the time).

6.3. *Patterns of Planet Occurrence*

Figure 4 shows graphically some of the key features of close-in planet occurrence. The number of planets per star varies by three orders of magnitude in the radius-period plane (Figure 4) that spans periods less than 50 days and planet radii less than 32 $R_\oplus$. Planet occurrence increases toward smaller radii (see Figure 5) down to our completeness limit of 2 $R_\oplus$, with a power law dependence given by $df(R)/d\log_{10} R = k_R R^\alpha$ where $df(R)/d\log_{10} R$ is the number of planets per star, for planets with $P < 50$ days in a $\log_{10}$ radius interval centered on R (in $R_\oplus$), $k_R = 2.9^{+0.5}_{-0.4}$, and $\alpha = -1.92 \pm 0.11$. This is a remarkable result, showing that from planets larger than Jupiter to those only twice the radius

of Earth planet occurrence rises rapidly by nearly two orders of magnitude. This rise with smaller size is consistent with, and supports the measured rise of, the planet occurrence with decreasing planet *mass* found by Howard et al. (2010). The increased occurrence of small planets seen in both studies supports the core-accretion theory for planet formation (Pollack et al. 1996).

Planet occurrence also increases with orbital period (Figure 6) in equal intervals of $\log P$ as $f(P) = k_P P^\beta \left(1 - e^{-(P/P_0)^\gamma}\right)$ with coefficients that all depend on planet radius, and both β and γ being positive. This functional form traces the steep rise in planet occurrence near a cut-off period, P_0. Below P_0 planets are rare, but for longer periods the planet occurrence distribution rises modestly with a power law dependence. We find that P_0 and γ (which governs how steep the occurrence fall-off is below P_0) depend on planet radius. The smaller planets, $R_p = 2$–$4 R_\oplus$, have $P_0 \sim 7$ days, while larger planets have a $P_0 \sim 2$ days. Further, γ is larger for planets with $R_p > 4 R_\oplus$ making the fall-off in planet occurrence more abrupt below P_0. The trends suggest that the mechanisms that caused the planets to migrate and stop at close orbital distances depend on planet size. Alternatively, if a substantial number of small close-in planets formed by *in situ* accretion, then our measurements trace the contours of this process (Raymond et al. 2008).

This period dependence of planet occurrence seems to contradict the results from Doppler surveys of exoplanets for which we find a pile-up of planets at periods of 3 days and a nearly flat distribution of planets for longer periods, out to periods of 1 yr (Wright et al. 2009). The key difference is that *Kepler* is sensitive to much smaller planets (in radius and mass) than were Doppler surveys, especially beyond 0.1 AU. To be sure, *Kepler* suffers a geometrical decline in detectability as $R_\star/a$ but we have corrected for this trivially. Such a correction is more difficult for Doppler surveys that have less uniform detectability from star to star.

Another difference in the period distributions between *Kepler* and Doppler surveys is in the pile-up of hot Jupiters at orbital periods near 3 days (Figures 4 and 6). The *Kepler* detected planets show a pile-up, but it is modest, almost not significant, while for single planets in Doppler surveys the pile-up is a factor of three above the background occurrence at other periods (Marcy et al. 2008; Wright et al. 2009). This different planet occurrence for hot Jupiters appears to be real, and may be due to fewer metal-rich stars in the *Kepler* sample that are located 50–200 pc above the Galactic plane, or different stellar mass distributions in the magnitude-limited and volume-limited surveys. The *Kepler* field has a greater admixture of thick disk stars (that are metal poor with [Fe/H] ≈ -0.5) to thin disk stars than do the Doppler target stars. Other photometric surveys have noted that hot Jupiter occurrence appears to vary with stellar population. Gilliland et al. (2000) found no planets in a *Hubble Space Telescope* survey of the globular cluster 47 Tucanae and estimated a hot Jupiter occurrence that is an order of magnitude lower than in the solar neighborhood. Similarly, Weldrake et al. (2008) found no planets in the ω Centauri globular cluster and found the occurrence of hot Jupiters ($P = 1$–5 days) to be less than 0.0017 planets per star. Gould et al. (2006) found an occurrence of

$0.003^{+0.004}_{-0.002}$ hot Jupiters per star for $P = 3$–5 days, based on the magnitude-limited OGLE-III survey in the bulge of the Galaxy, and are compatible with our results from *Kepler*, 0.005 ± 0.001 planets per star for $R_p = 8$–32 $R_\oplus$, $P < 50$ days, and Kp < 16.

We further find that planets larger than 16 $R_\oplus$ (1.5 R_{Jup}) are extremely rare. Such inflated planets are also rare among transiting planets detected from the ground (see, e.g., the mass-radius diagram for gas giant planets in Bakos et al. (2010b)). For several Gyr-old planets, theoretical mass-radius curves show a maximum near $R_p \approx 13 R_\oplus \approx 1.2 R_{Jup}$ (Fortney et al. 2007b). Larger planets are typically young or close-in and inflated by one of several proposed mechanisms (e.g., Batygin & Stevenson 2010; Laughlin et al. 2011; Burrows et al. 2007).

We also note some interesting morphology in the two-dimensional occurrence domain of planet radius and orbital period (Figure 4). There is a ridge of higher planet occurrence for super-Earths and Neptunes, similar to that identified in Howard et al. (2010). The ridge appears to be diagonal when plotting either M_p or R_p vs. P extending from a period and radius of 3 days and 2 $R_\oplus$ (lower left) to a period and radius of 50 days and 4 $R_\oplus$. This ridge can be seen by direct inspection of the Figure 4, both by the density of the dots and by the colors. The upper envelope of red boxes (indicating high planet occurrence) extends along a diagonal from lower left to upper right. This ridge conveys some key information about the formation and perhaps dynamical evolution or migration of the 2–4 $R_\oplus$ planets.

The paucity of close-in Neptune-mass planets ($M_p \sin i = 10$–100 $M_\oplus$, $P < 20$ days) seen in Howard et al. (2010) is not as clearly visible in the Kepler data. In particular, the "top" of this desert ($M_p \sin i = 100 M_\oplus$, or the radius equivalent) is not as clear. A further study of *Kepler* stars to fainter magnitudes of Kp = 16 may shed light on this desert. The overall planet occurrence for GK stars and periods less than 50 days, listed in Table 3, shows that planets of 2–4 $R_\oplus$ is 0.130 ± 0.008 planets per star. This agrees well with the planet occurrence of 3–30 $M_\oplus$ planets found by Howard et al. (2010) of 15^{+5}_{-4}%. The planet occurrence for all planet radii from 2–32 $R_\oplus$ is only 16.5%, again in agreement with Howard et al. (2010) and Cumming et al. (2008). We find little support for the suggestion of planet occurrences of super-Earths and Neptunes ($M_p \sin i = 3$–30 $M_\oplus$) of 30% $\pm$ 10% (Mayor et al. 2009) for $P < 50$ days.

We also measured planet occurrence as a function of T_{eff} of the host star, a proxy for stellar mass. For the smallest planets, 2–4 $R_\oplus$, the results show a nearly linear rise in planet occurrence with smaller stellar mass. One may wonder if this rise might be caused by some systematic error due to poor values of T_{eff} or $R_\star$ in the *Kepler* Input Catalog. Such a systematic error seems nearly impossible, as the KIC values of T_{eff} are accurate to 135 K (RMS) and in any case the T_{eff} values certainly vary monotonically with the true value of T_{eff} even if one imagines some large systematic error in the KIC values of T_{eff}. Thus the increase in planet occurrence with smaller T_{eff} and hence smaller stellar mass, appears to be real. Again, we emphasize that the SNR = 10 criterion for a *Kepler* target star to be included in our survey implies

that the detection efficiency is close to unity for all stars, from 7100 K to 3600 K, for $R > 2$ $R_\oplus$. Examination of Figure 8 shows that even if one ignores the coolest and hottest stars, the increase of planet occurrence persists robustly. Thus it appears that the number of planets per star increases by a factor of seven from stars of 1.5 $M_\odot$ to stars of 0.4 $M_\odot$ ($T_{\rm eff} = 7100$–3600K), with all of that $T_{\rm eff}$ dependence coming from the smallest planets, 2-4 $R_\oplus$. This high occurrence of close-in small planets around low mass stars represents significant information about the formation mechanisms of planets of 2–4 $R_\oplus$.

We considered the possibility that this correlation is due to a systematic metallicity bias that depends on $T_{\rm eff}$. That is, cool stars are relatively nearby, close to the galactic plane, and have higher metallicities, while hot stars are on average more distant, at greater heights above the galactic plane, and have lower metallicities. In this scenario, low metallicity is the driving force behind lower planet occurrence at higher $T_{\rm eff}$. Using the Besancon galactic model, we estimate that metallicities may vary by ~0.1 dex as a function of $T_{\rm eff}$, but the dependence need not be monotonic because of the median age varies with $T_{\rm eff}$. It would be remarkable if such a modest difference in metallicity could cause a factor of seven difference in close-in planet occurrence. Unfortunately, due to the poor [Fe/H] measurements in the KIC, we are unable to measure the occurrence of planets as a function of [Fe/H]. Note, however, that either result has profound implications for planet formation: the occurrence of 2–4 $R_\oplus$ planets depends strongly on stellar properties, $T_{\rm eff}$ or [Fe/H].

Sub-Neptune-size and jovian planets appear to have opposite trends in occurrence as a function of $M_\star$. We showed that the occurrence of 2–4 $R_\oplus$ planets decreases by a factor of seven with $M_\star$ over ~0.4–1.5 $M_\odot$ ($T_{\rm eff} = 3600$–7100 K). Johnson et al. (2010) measured the occurrence of giant planets as a function of $M_\star$ and [Fe/H] and found a positive correlation with both quantities. That is, the occurrence of giant planets *increases* with increasing $M_\star$ over the range ~0.3–1.9 $M_\odot$. Their study considered only giant planets that produce $K > 20$ m s^{-1} Doppler signals and orbit within 2.5 AU. Sub-giants with $M_\star = 1.4$–1.9 $M_\odot$ have the highest rate of giant planet occurrence in their study. However, most of these planets orbit at ~1–2 AU, with almost no planets inside of $P = 50$ day orbits (Bowler et al. 2010). Close-in planets of all sizes larger than 2 $R_\oplus$ appear to be rare around the most massive stars accessible to transit and Doppler surveys.

6.4. *Planet Formation*

Population synthesis models of planet formation by core accretion simulate the growth and migration of planet embryos embedded in a proto-planetary disk of gas and dust. Among their key predictions is the distribution of planet mass or radius as a function of orbital distance. Early versions of these models (Ida & Lin 2004; Alibert et al. 2005; Mordasini et al. 2009a) were tuned to match the distribution of giant planets detected by RV (Cumming et al. 2008; Udry & Santos 2007) by decreasing the rate of Type I migration compared to theoretical predictions (Ward 1997; Tanaka et al. 2002). The simulations predicted that planet occurrence rises with decreasing planet mass. But most of the low-mass plan-

ets resided in orbits near or beyond the ice line at ~2–3 AU. These models also robustly predicted a "planet desert", a region of parameter space nearly devoid of planets. Planets with $M_{\rm p} \sin i \approx 1$–20 $M_\oplus$ and $a \lesssim 1$ AU were predicted to be extremely rare because producing such planets requires the gas disk to dissipate while one of two faster processes were happening, Type II migration or run-away gas accretion. Meanwhile, the models predicted that planets with masses above the desert, $M > 20 M_\oplus$, but residing inside of ~1 AU would exhibit a nearly constant distribution with mass.

Howard et al. (2010) demonstrated that the observed distribution of close-in planets ($P < 50$ days) exhibited quite different properties from those predicted by population synthesis. The predicted planet desert is actually populated by the highest planet occurrence of any region of the mass-period parameter space yet probed (the "ridge" noted above). The planet mass function rises steeply with decreasing planet mass, in contradiction to the expected nearly constant distribution with mass outside of the desert. From *Kepler*, we also see many planets populating the predicted desert (Figure 4) and a planet radius distribution that rises steeply with decreasing planet size (tracking the mass distribution). The latest versions of the population synthesis models (Ida & Lin 2010; Alibert et al. 2011) offer improvements including non-isothermal treatment of the disk (Paardekooper et al. 2010) and multiple, interacting planet embryos per simulation. But they still predict a planet desert (albeit partially filled in). The contours of planet occurrence in Figure 4 offer rich detail to which future refinements of these models can be tuned. Alternatively, the distribution of observed planets may be strongly shaped by processes that take place after the gas clears, namely planet-planet scattering (e.g., Ford et al. 2005; Ford & Rasio 2008; Chatterjee et al. 2008; Raymond et al. 2009), secular and resonant migration (e.g., Lithwick & Wu 2010; Wu & Lithwick 2011), and planetesimal migration and growth (e.g., Kirsh et al. 2009; Capobianco et al. 2011; Walsh & Morbidelli 2011). If these processes strongly shape the final planet distributions, then the planet distributions from population synthesis models (which truncate when the gas clears) will form the input to additional simulations that model post-disk effects and hope to match the presently observed planet distributions.

Current planet formation theory must also adapt to account for remarkable orbital properties of exoplanets. Not included here is an analysis of the orbital eccentricities that span the range $e = 0$–0.93 (e.g., Marcy et al. 2005b; Udry & Santos 2007; Moorhead et al. 2011) and the close-in "hot Jupiters" show a wide distribution of inclinations relative to the equatorial plane of the host star (Winn et al. 2010, 2011; Triaud et al. 2010; Morton & Johnson 2010). Thus, standard planet formation theory probably requires additional planet-planet gravitational interactions to explain these non-circular and non-coplanar orbits (Chatterjee et al. 2010; Wu & Lithwick 2011).

6.5. *The Future of Kepler*

We strongly advocate for an improved catalog of stellar parameters for the ~1000 *Kepler* planet host stars and a comparably-sized control sample. Our occurrence mea-

surements and their interpretations would be strengthened by an improved knowledge of $R_\star$, $\log g$, [Fe/H], and $T_{\rm eff}$. The $R_\star$ values from the KIC are only known to 35% (rms) which leads to a proportionally large uncertainty in $R_{\rm p}$. We saw that hot Jupiters have a significantly lower occurrence in the *Kepler* sample than in RV surveys. We were unable to test whether this is due to differing metallicities of the host stars because [Fe/H] is poorly measured in the KIC. Similarly, we are unable to completely rule out a metallicity gradient with height above the galactic plane as the underlying cause of the observed seven-fold decrease in the occurrence of 2–4 $R_\oplus$ planets with increasing $T_{\rm eff}$.

Finally, we note that Figure 3 shows representative planets having $R_{\rm p} \sim 2.5$ $R_\oplus$ and $P < 50$ days, all of which reach SNR ~ 20 in four quarters of *Kepler* photometry (and SNR ~ 10 in one quarter). If we consider the SNR for planets of radius 1 $R_\oplus$, the transit depth is 6 times shallower, implying total SNR values near SNR = 20 / 6 = 3.3. Thus, planets of 1 $R_\oplus$, even in short periods under 50 days, would not reach the threshold SNR for meriting a secure detection with current data in hand. For planets of 1 $R_\oplus$ to reach SNR ~ 6.6, *Kepler* must acquire four times more data, i.e. five years total, still constituting a marginal detection. Clearly an extended mission of an additional ~ 3 yr is needed to bring planets of 1 $R_\oplus$ to SNR > 7.

REFERENCES

Adams, E. R., Seager, S., & Elkins-Tanton, L. 2008, ApJ, 673, 1160

Alibert, Y., Mordasini, C., & Benz, W. 2011, A&A, 526, A63+

Alibert, Y., Mordasini, C., Benz, W., & Winisdoerffer, C. 2005, A&A, 434, 343

Bakos, G. Á., et al. 2010a, ApJ, 710, 1724

—. 2010b, ApJ submitted, arXiv:1008.3388

Baraffe, I., Chabrier, G., & Barman, T. 2008, A&A, 482, 315

Batalha, N. M., et al. 2010, ApJ, 713, L109

—. 2011, ApJ, 729, 27

Batygin, K., & Stevenson, D. J. 2010, ApJ, 714, L238

Bensby, T., Zenn, A. R., Oey, M. S., & Feltzing, S. 2007, ApJ, 663, L13

Bordé, P., et al. 2010, A&A, 520, A66+

Borucki, W. J., et al. 2010a, ApJ submitted, arXiv:1006.2799

—. 2010b, ApJ, 713, L126

—. 2011, ApJ accepted, arXiv:1102.0541

Bowler, B. P., et al. 2010, ApJ, 709, 396

Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, AJ submitted, arXiv:1102.0342

Burrows, A., Hubeny, I., Budaj, J., & Hubbard, W. B. 2007, ApJ, 661, 502

Caldwell, D. A., et al. 2010, ApJ, 713, L92

Capobianco, C. C., Duncan, M., & Levison, H. F. 2011, Icarus, 211, 819

Chatterjee, S., Ford, E. B., Matsumura, S., & Rasio, F. A. 2008, ApJ, 686, 580

Chatterjee, S., Ford, E. B., & Rasio, F. A. 2010, arXiv:1012.0584

Cox, A. N., ed. 2000, Allen's Astrophysical Quantities (Springer)

Cumming, A., Butler, R. P., Marcy, G. W., Vogt, S. S., Wright, J. T., & Fischer, D. A. 2008, PASP, 120, 531

Dunham, E. W., et al. 2010, ApJ, 713, L136

Ferraz-Mello, S., Tadeu dos Santos, M., Beauge, C., Michtchenko, T. A., & Rodriguez, A. 2010, arXiv:1011.2144

Fischer, D. A., & Valenti, J. 2005, ApJ, 622, 1102

Ford, E. B., Lystad, V., & Rasio, F. A. 2005, Nature, 434, 873

Ford, E. B., & Rasio, F. A. 2008, ApJ, 686, 621

Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007a, ApJ, 668, 1267

—. 2007b, ApJ, 659, 1661

Gautier, III, T. N., et al. 2010, ApJ submitted, arXiv:1001.0352

Gilliland, R. L., et al. 2000, ApJ, 545, L47

Gillon, M., et al. 2007, A&A, 472, L13

Gould, A., Dorsher, S., Gaudi, B. S., & Udalski, A. 2006, Acta Astronomica, 56, 1

Grasset, O., Schneider, J., & Sotin, C. 2009, ApJ, 693, 722

Hartman, J. D., et al. 2011, ApJ, 728, 138

Hatzes, A. P., et al. 2010, A&A, 520, A93+

Holman, M. J., et al. 2010, Science, 330, 51

Howard, A. W., et al. 2009, ApJ, 696, 75

—. 2010, Science, 330, 653

—. 2011a, ApJ, 726, 73

—. 2011b, ApJ, 730, 10

Ida, S., & Lin, D. N. C. 2004, ApJ, 604, 388

—. 2008a, ApJ, 673, 487

—. 2008b, ApJ, 685, 584

—. 2010, ApJ, 719, 810

Jenkins, J. M., et al. 2010a, ApJ, 724, 1108

—. 2010b, ApJ, 713, L120

—. 2010c, ApJ, 713, L87

Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010, PASP, 122, 905

Johnson, J. A., Winn, J. N., Albrecht, S., Howard, A. W., Marcy, G. W., & Gazak, J. Z. 2009, PASP, 121, 1104

Kepler Mission Team. 2009, VizieR Online Data Catalog, 5133

Kirsh, D. R., Duncan, M., Brasser, R., & Levison, H. F. 2009, Icarus, 199, 197

Koch, D. G., et al. 2010a, ApJ, 713, L131

—. 2010b, ApJ, 713, L79

Latham, D. W., et al. 2010, ApJ, 713, L140

Laughlin, G., Crismani, M., & Adams, F. C. 2011, ApJ, 729, L7+

Léger, A., et al. 2009, A&A, 506, 287

Lissauer, J. J., Hubickyj, O., D'Angelo, G., & Bodenheimer, P. 2009, Icarus, 199, 338

Lissauer, J. J., et al. 2011a, Nature, 470, 53

—. 2011b, arXiv:1102.0543

Lithwick, Y., & Wu, Y. 2010, arXiv:1012.3706

Maness, H. L., Marcy, G. W., Ford, E. B., Hauschildt, P. H., Shreve, A. T., Basri, G. B., Butler, R. P., & Vogt, S. S. 2007, PASP, 119, 90

Marcy, G., Butler, R. P., Fischer, D., Vogt, S., Wright, J. T., Tinney, C. G., & Jones, H. R. A. 2005a, Progress of Theoretical Physics Supplement, 158, 24

Marcy, G. W., Butler, R. P., Vogt, S. S., Fischer, D. A., Henry, G. W., Laughlin, G., Wright, J. T., & Johnson, J. A. 2005b, ApJ, 619, 570

Marcy, G. W., et al. 2008, Physica Scripta Volume T, 130, 014001

Mayor, M., et al. 2009, A&A, 493, 639

Moorhead, A. V., et al. 2011, arXiv:1102.0547

Mordasini, C., Alibert, Y., Benz, W., & Naef, D. 2009a, A&A, 501, 1161

—. 2009b, A&A, 501, 1161

Morton, T. D., & Johnson, J. A. 2010, ApJ submitted, arXiv:1010.4025

—. 2011, ApJ submitted, arXiv:1101.5630

Movshovitz, N., Bodenheimer, P., Podolak, M., & Lissauer, J. J. 2010, Icarus, 209, 616

Nagasawa, M., Ida, S., & Bessho, T. 2008, ApJ, 678, 498

Nettelmann, N., Fortney, J. J., Kramm, U., & Redmer, R. 2010, ApJ submitted, arXiv:1010.0277

Neves, V., Santos, N. C., Sousa, S. G., Correia, A. C. M., & Israelian, G. 2009, A&A, 497, 563

O'Donovan, F. T., et al. 2006, ApJ, 651, L61

Paardekooper, S., Baruteau, C., Crida, A., & Kley, W. 2010, MNRAS, 401, 1950

Pál, A., et al. 2008, ApJ, 680, 1450

Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., & Greenzweig, Y. 1996, Icarus, 124, 62

Pont, F., Aigrain, S., & Zucker, S. 2011, MNRAS, 411, 1953

Queloz, D., et al. 2009, A&A, 506, 303

Raymond, S. N., Barnes, R., & Mandell, A. M. 2008, MNRAS, 384, 663

Raymond, S. N., Barnes, R., Veras, D., Armitage, P. J., Gorelick, N., & Greenberg, R. 2009, ApJ, 696, L98

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

Rowe, J. F., et al. 2010, ApJ, 713, L150

Sato, B., et al. 2005, ApJ, 633, 465

Schlaufman, K. C., Lin, D. N. C., & Ida, S. 2009, ApJ, 691, 1322

—. 2010, ApJ, 724, L53

Seager, S., Kuchner, M., Hier-Majumder, C. A., & Militzer, B. 2007, ApJ, 669, 1279

Sotin, C., Grasset, O., & Mocquet, A. 2007, Icarus, 191, 337

Tanaka, H., Takeuchi, T., & Ward, W. R. 2002, ApJ, 565, 1257

Torres, G., Winn, J. N., & Holman, M. J. 2008, ApJ, 677, 1324

Torres, G., et al. 2011, ApJ, 727, 24

Triaud, A. H. M. J., et al. 2010, A&A, 524, A25+

Udry, S., & Santos, N. C. 2007, ARA&A, 45, 397

Valencia, D., O'Connell, R. J., & Sasselov, D. 2006, Icarus, 181, 545

Valenti, J. A. 2010, in IAU Symposium, Vol. 265, IAU Symposium, ed. K. Cunha, M. Spite, & B. Barbuy, 403–407

Valenti, J. A., & Fischer, D. A. 2005, ApJS, 159, 141

von Braun, K., Kane, S. R., & Ciardi, D. R. 2009, ApJ, 702, 779

Walsh, K. J., & Morbidelli, A. 2011, A&A, 526, A126+

Ward, W. R. 1997, ApJ, 482, L211+

Weldrake, D. T. F., Sackett, P. D., & Bridges, T. J. 2008, ApJ, 674, 1117

Winn, J. N., Fabrycky, D., Albrecht, S., & Johnson, J. A. 2010, ApJ, 718, L145

Winn, J. N., et al. 2011, AJ, 141, 63

Wright, J. T., Upadhyay, S., Marcy, G. W., Fischer, D. A., Ford, E. B., & Johnson, J. A. 2009, ApJ, 693, 1084

Wu, Y., & Lithwick, Y. 2011, ApJ submitted, arXiv:1012.3475