

ON THE OCCURRENCE RATE OF HOT JUPITERS IN DIFFERENT STELLAR ENVIRONMENTS

Ji WANG¹, DEBRA A. FISCHER¹, ELLIOTT P. HORCH², AND XU HUANG³

¹ Department of Astronomy, Yale University, New Haven, CT 06511, USA; ji.wang@yale.edu

² Department of Physics, Southern Connecticut State University, 501 Crescent Street, New Haven, CT 06515, USA

³ Department of Astrophysical Sciences, Peyton Hall, 4 Ivy Lane, Princeton University, Princeton, NJ 08540, USA

Received 2014 November 18; accepted 2014 December 4; published 2015 January 30

ABSTRACT

Many hot Jupiters (HJs) are detected by the Doppler and transit techniques. From surveys using these two techniques, however, the measured HJ occurrence rates differ by a factor of two or more. Using the California Planet Survey sample and the *Kepler* sample, we investigate the causes for this difference in the HJ occurrence rate. First, we find that $12.8\% \pm 0.24\%$ of HJs are misidentified in the *Kepler* mission because of photometric dilution and subgiant contamination. Second, we explore the differences between the Doppler sample and the *Kepler* sample that can account for the different HJ occurrence rate. Third, we discuss how to measure the fundamental HJ occurrence rates by synthesizing the results from the Doppler and *Kepler* surveys. The fundamental HJ occurrence rates are measures of the HJ occurrence rate as a function of stellar multiplicity and evolutionary stage, e.g., the HJ occurrence rate for single and multiple stars or for main-sequence and subgiant stars. While we find qualitative evidence that HJs occur less frequently in subgiants and multiple stellar systems, we conclude that our current knowledge of stellar properties and the stellar multiplicity rate is too limited for us to reach any quantitative result for the fundamental HJ occurrence rates. This concern extends to η_{Earth} , the occurrence rate of Earth-like planets.

Key words: methods: statistical – planets and satellites: fundamental parameters – stars: solar-type – stars: statistics – techniques: photometric – techniques: radial velocities

1. INTRODUCTION

Hot Jupiters (HJs) are among the most prominent astronomical discoveries of the past century (Mayor & Queloz 1995; Marcy & Butler 1996). Their existence challenged the previously accepted classic planet formation model, which is tailored for the solar system (Lissauer 1993). Our knowledge of planet formation has moved beyond the solar system ever since the first HJ discovery.

Among all exoplanets, HJs are the easiest to detect by the Doppler technique because of their short periods ($P < 10$ days) and large planetary masses ($m > 0.1 M_J$). Doppler planet surveys show that the occurrence rate of HJs is $\sim 1.0\%$, though the exact numbers slightly differ for different surveys. For example, Wright et al. (2012) estimated that the occurrence rate of HJs is $1.2\% \pm 0.38\%$ based on data obtained from the Keck and Lick Observatories. Other studies based on a similar set of data found the HJ occurrence rate to be $1.2\% \pm 0.1\%$ (Marcy et al. 2005) and $1.5\% \pm 0.6\%$ (Cumming et al. 2008). Based on radial velocity (RV) data obtained from the HARPS and ELODIE, Mayor et al. (2011) measured the HJ occurrence rate at $0.89\% \pm 0.36\%$.

The HJ occurrence rate measured from the Doppler planet surveys is at odds with measurements from transit surveys. In the Optical Galactic Lensing Experiment III survey, Gould et al. (2006) estimated that the HJ occurrence rate is $0.31^{+0.43}_{-0.18}\%$. The number from the SuperLupus Survey is $0.10^{+0.27}_{-0.08}\%$ (Bayliss & Sackett 2011). A more recent result from the *Kepler* mission suggested that the HJ occurrence rate is $0.4\% \pm 0.1\%$ ($8.0 R_{\oplus} \leq R_p \leq 32.0 R_{\oplus}$, $P < 10$ days; Howard et al. 2012) or $0.43\% \pm 0.05\%$ ($6.0 R_{\oplus} \leq R_p \leq 22.0 R_{\oplus}$, $0.8 \text{ days} \leq P \leq 10.0$ days; Fressin et al. 2013). These numbers are a factor two to three smaller than those from Doppler surveys.

Potential explanations of the difference include stellar metallicity, age, and population (Howard et al. 2012; Wright et al. 2012; Gould et al. 2006; Bayliss & Sackett 2011). Stars with

higher metallicity are more likely to host HJs (Gonzalez 1997; Santos et al. 2001; Fischer & Valenti 2005; Sousa et al. 2011; Wang & Fischer 2013). For example, at least 25% of stars with twice the solar metallicity are orbited by a giant planet, and this number drops to less than 5% for stars with solar metallicity (Udry & Santos 2007). In the sub-solar metallicity regime, the occurrence rate of HJ is found to be below 1% at $0.37^{+0.6}_{-0.4}\%$ (Mortier et al. 2012). If the stellar metallicity for stars in Doppler surveys is systematically higher than those in transit surveys, then the HJ occurrence rate from the Doppler surveys is higher than it is for the transit surveys. Given the strong dependence of the HJ occurrence rate on stellar metallicity (Fischer & Valenti 2005), a metallicity difference of 0.15 dex may explain the difference of HJ occurrence rate.

Stellar age may also explain the difference in the HJ occurrence rate. If *Kepler* stars are older on average, then the fraction of evolved stars is higher. Since there is evidence that evolved stars host fewer HJs (Bowler et al. 2010; Johnson et al. 2010; Schlaufman & Winn 2013), a higher fraction of evolved stars may lead to a lower HJ occurrence rate for the *Kepler* sample.

Another possible explanation may be stellar population. Transit surveys usually target the galactic bulge (e.g., Gould et al. 2006; Bayliss & Sackett 2011), where the stellar population is dominated by M dwarfs (Henry et al. 1997). In comparison, Doppler surveys usually selected solar-type stars (e.g., Valenti & Fischer 2005; Sousa et al. 2011). Since HJs occur less frequently around M dwarfs (Johnson et al. 2010; Bonfils et al. 2013), a sample of stars with a higher fraction of M dwarfs leads to a lower HJ occurrence rate. However, the stellar population difference is less of a concern for the comparison between the *Kepler* sample and the Doppler sample. Unlike other transit surveys, *Kepler* stars mainly consist of solar-type stars (Brown et al. 2011).

In this paper, we investigate other possibilities that may reconcile the difference in the HJ occurrence rates from the *Kepler* and Doppler surveys. In Section 2, we quantify the

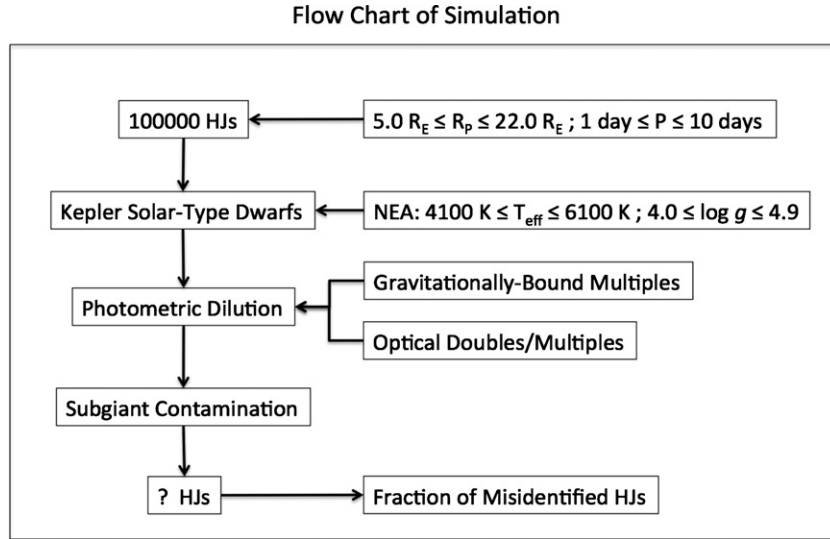


Figure 1. Flow chart of a simulation to quantify the fraction of HJs that are misidentified as smaller planets due to two effects: photometric dilution and subgiant contamination.

fraction of HJs that are misidentified from the *Kepler* mission because of photometric dilution and subgiant contamination. In Section 3, we summarize the potential explanations for the discrepancy of HJ occurrence rate. In Sections 4 and 5, we synthesize the *Kepler* and Doppler results and discuss how to probe the fundamental HJ occurrence rate as a function of the stellar multiplicity rate and evolutionary stage. In Section 6, we discuss our conclusions and their implications to future investigations of planet occurrence rate.

2. FRACTION OF MISIDENTIFIED HJs

2.1. Simulated HJs

We attempt to quantify the fraction of HJs that are misidentified as smaller planets due to two effects: photometric dilution and subgiant contamination. Figure 1 shows the flow chart of our simulation. We start with 100,000 HJs ($1 \text{ day} \leq P \leq 10 \text{ days}$, $5.0 R_{\oplus} \leq R_p \leq 22.0 R_{\oplus}$). The lower limit of HJs roughly corresponds to $0.1 M_J$ assuming the mass–radius relationship used in Lissauer et al. (2011). The radius distribution of these HJs follows a power law of -2.92 ± 0.11 (Howard et al. 2012). The period distribution between 2 and 10 days follows a power law of -1 , i.e., uniform distribution in logarithmic space. The period distribution between one and two days follows the same power-law distribution but with half the probability of $2 \text{ day} \leq P \leq 10 \text{ days}$. This assumption accounts for the lack of HJs within the two-day orbital period (Howard et al. 2012).

2.2. Kepler Solar-type Dwarf Stars

We obtain *Kepler* stellar properties from the NASA Exoplanet Archive (NEA; Huber et al. 2014).⁴ We apply a cut in effective temperature ($4100 \text{ K} \leq T_{\text{eff}} \leq 6100 \text{ K}$) and surface gravity ($4.0 \leq \log g \leq 4.9$), which results in 109,335 solar-type dwarf stars in the *Kepler* sample. The cut is consistent with the solar-type star definition from Howard et al. (2012) for a proper comparison to previous result.

2.3. Photometric Dilution in Multiple Stellar Systems

The transit signal of a planet is diluted by the flux of companion stars. The dilution leads to two consequences: (1)

a missing planet or (2) a planet with an underestimated planet radius in the detection case. For the *Kepler* mission, a planet is missed when the detection signal-to-noise ratio (S/N) is lower than 7.1 (Jenkins et al. 2010). The S/N is calculated based on the following equation (Wang et al. 2014a):

$$S/N = \frac{\delta}{\text{CDPP}_{\text{eff}}} \sqrt{N_{\text{transits}}}, \quad (1)$$

where δ is transit depth, CDPP_{eff} is the effective combined differential photometric precision (Jenkins et al. 2010), a measure of photometric noise, and N_{transits} is the number of observed transits. The transit depth is calculated by the following equation:

$$\delta = \frac{R_{\text{PL}}^2}{R_*^2} \frac{F_*}{F_* + F_c}, \quad (2)$$

where R_{PL} is planet radius, R_* is the radius of the star that the planet is transiting, F denotes flux, and subscripts $*$ and c indicate the planet host star and the contaminating star, respectively. Throughout the paper, we only consider the case in which an HJ transits the primary star because we focus on HJs around solar-type stars and the secondary star may not be a solar-type star in most cases.

Even if a planet is detected in the presence of photometric dilution, the photometric dilution will cause some HJs to be misidentified as smaller planets. These misidentified HJs are not accounted for in the statistics for HJs. Thus, the fraction of misidentified HJs needs to be quantified and corrected for. The measured planet radius (R'_{PL}) can be calculated by the following equation:

$$R'_{\text{PL}} = R_{\text{PL}} \sqrt{\frac{F_*}{F_* + F_c}}. \quad (3)$$

2.3.1. Gravitationally Bound Multiple Stellar Systems

Companion stars may be gravitationally bound or optical doubles or multiples (i.e., unbound). We consider these two cases separately. For gravitationally bound multiple stellar systems, the stellar multiplicity rate for planet host stars (Wang et al. 2014a, 2014b) is significantly lower than that for the solar neighborhood (Raghavan et al. 2010). We adopt the stellar multiplicity rate measured from Wang et al. (2014a)

⁴ <http://exoplanetarchive.ipac.caltech.edu/>

for stellar separations smaller than 1000 AU, i.e., $24\% \pm 7\%$. Beyond 1000 AU, we adopt the log-normal distribution of stellar separation from Duquennoy & Mayor (1991). At these separations, there is no significant difference of stellar multiplicity between Wang et al. (2014a) and Duquennoy & Mayor (1991). In addition, the result from Wang et al. (2014a) beyond 1000 AU is prone to contamination of optical doubles and multiples, which will be addressed in the following section. In our simulation for gravitationally bound systems, the mass ratio of stellar companions to the primary star follows a normal distribution with a mean and standard deviation of 0.23 and 0.42 (Duquennoy & Mayor 1991). The conversion between stellar mass and radius is based on Feiden & Chaboyer (2012). The conversion from stellar mass to stellar flux in the *Kepler* band is based on Kraus & Hillenbrand (2007).

2.3.2. Optical Doubles and Multiples

For photometric dilution caused by optical doubles and multiples, we use the TRILEGAL galaxy model (Girardi et al. 2005) to study the probability of such cases. Horch et al. (2014) describes the process in detail. Ten fields with a field of view of 1 deg^2 are simulated. These fields have different galactic latitudes, thus the combination of the results from the fields gives a better statistical result of the entire *Kepler* field of view. In each of field, the TRILEGAL model is used to construct a simulated stellar population with effective temperature distribution that is similar to *Kepler* stars. Binary parameters are turned off in the simulation because we focus on optical doubles and multiples, gravitationally bound multiple stellar systems have already been discussed.

From the simulations, we calculated the probability of optical doubles or multiples. We find that 55.7% of the simulated stars have at least one visual companion within $4''$ down to the *Kepler* magnitude of 27.0. The separation of $4''$ corresponds to the pixel scale of the *Kepler* CCD detector. Since stellar companions with small differential magnitudes cause a large adjustment of planet radius (Equation (3)), we also report the fraction of simulated stars with stellar companions down to a certain differential magnitude. For differential magnitudes of 1, 3, and 5 mag, the fractions of simulated stars with stellar companions are 2.8%, 11.3%, and 26.7%, respectively. Figure 2 shows a contour plot of joint probability of $Kmag_1$ and $Kmag_2$, where $Kmag_1$ is the *Kepler* magnitude of the primary star and $Kmag_2$ is the *Kepler* magnitude of the secondary star. Optical doubles and multiples with small delta magnitude are more likely for faint *Kepler* stars ($Kmag_1 \geq 13.0$) than for bright *Kepler* stars ($Kmag_1 \leq 13.0$). In a transit observation, measured planet radius is affected more by brighter stellar companions (see Equation (3)), i.e., secondary star with smaller delta magnitude, and thus misidentified HJs are more likely to take place for faint *Kepler* stars.

2.4. Contamination of Subgiants

In a transit observation, the ratio of planet radius to star radius is measured. When converting the ratio to planet radius, the star radius needs to be multiplied. A systematic error in star radius estimation can lead to inaccurate planet radius and thus misidentify a planet into an incorrect category. While we use the NEA stellar property catalog to select solar-type dwarf stars, chances are that some the selected stars are actually subgiant stars. Reported planet radii from the NEA may be systematically lower than they should be with the

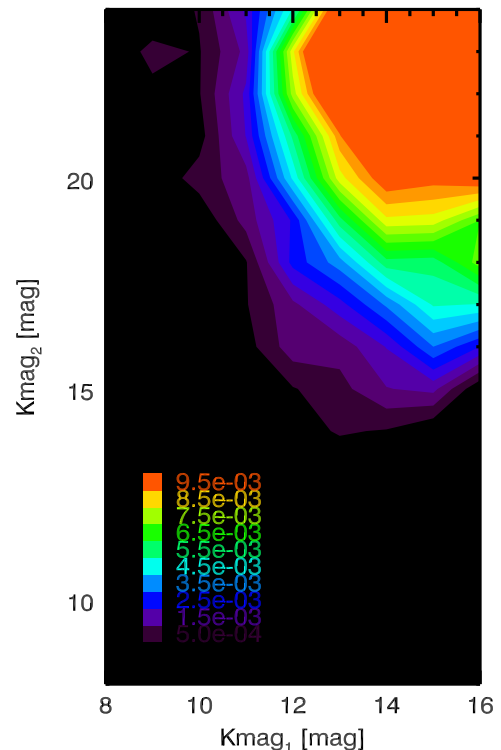


Figure 2. Contours of the joint probability distribution of $Kmag_1$ and $Kmag_2$ for optical doubles and multiples in the *Kepler* field of view. $Kmag_1$ is the *Kepler* magnitude of the primary star and $Kmag_2$ is the *Kepler* magnitude of the secondary star. 55.7% of *Kepler* stars have at least one visual unbound stellar companion. Fainter *Kepler* stars are more likely to have visual unbound stellar companions with small delta magnitudes. Therefore, planets around faint *Kepler* stars are more likely to be misidentified as smaller planets.

contamination of subgiants. Bastien et al. (2014) used a short-timescale photometric variation of bright *Kepler* stars ($Kmag \leq 13.0$) to infer stellar surface gravity and radius. They found that 48% of *Kepler* stars are modestly evolved subgiant stars. In comparison, 27% of *Kepler* stars are subgiants from the NEA statistics. The comparison indicates that 29% of *Kepler* “dwarf” stars according to the NEA are actually subgiants. These subgiants are, on average, 20%–30% larger than reported in the NEA. Therefore, radii of $\sim 30\%$ of *Kepler* planets are underestimated by 20%–30%. We use a value of 25% for the percentage by which the stellar radius is underestimated. To account for subgiant contamination in the *Kepler* dwarf sample, we randomly assign 30% of simulated stars with HJs as subgiants that are misidentified as “dwarf” stars. We adopt the adjusted subgiant radius to calculate S/N based on Equation (1). If the S/N is lower than 7.1, then the HJ is marked as a missing HJ due to insufficient S/N. In the case of detection ($S/N \geq 7.1$), the radius of the HJ is recorded with a smaller value by a factor of 1.25.

2.5. Result

We repeat the above simulation 100 times and calculate the fraction of HJs that are misidentified as smaller planets. We find that $12.48\% \pm 0.24\%$ of HJs are misidentified due to photometric dilution and subgiant contamination. Specifically, $2.11\% \pm 0.08\%$ and $1.17\% \pm 0.06\%$ of misidentifications are caused by photometric dilution due to gravitationally bound multiple stellar systems and optical doubles/multiples, respectively. $9.59\% \pm 0.15\%$ of misidentifications are caused by subgiant contamination. Therefore, subgiant contamination is the

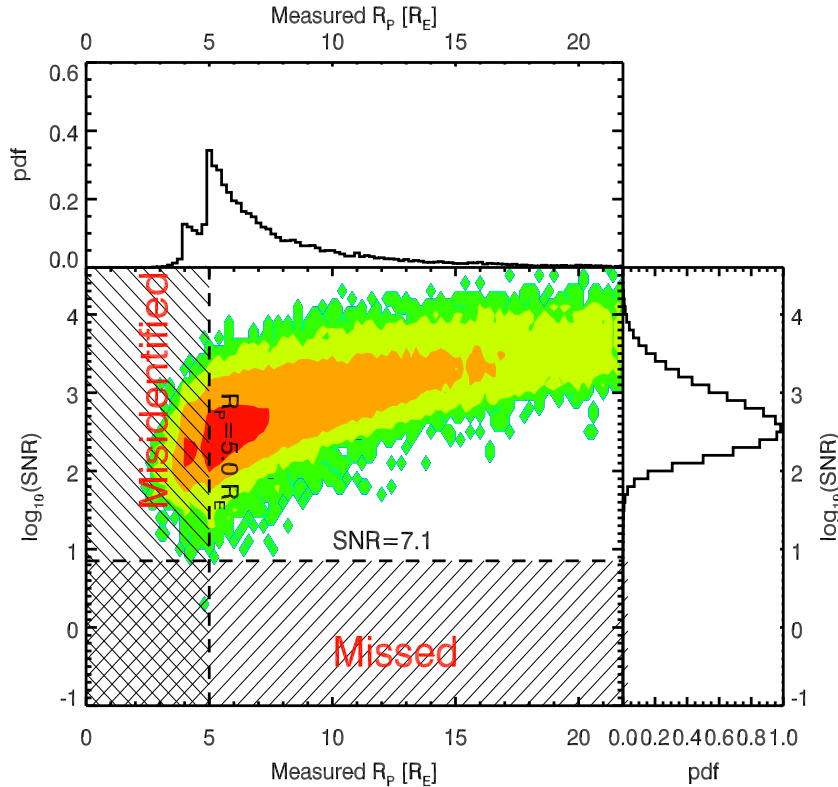


Figure 3. Result of the simulation to quantify the fraction of HJs that were misidentified because of photometric dilution and subgiant contamination. Color contours show the joint distribution of S/Ns and measured planet radius of simulated HJs. The marginalized distributions of S/Ns and measured planet radius are shown on the side of each axis. We find that $12.48\% \pm 0.24\%$ of HJs are misidentified due to photometric dilution and subgiant contamination.

major cause that is responsible for HJ misidentification. Figure 3 shows a contour plot for the distribution of HJ planet radius and S/N after considering the effects of photometric dilution and subgiant contamination. Because of large signal of HJs, they are rarely missed due to an insufficient S/N. The fraction of missing HJs is always less than 0.005% and is thus negligible compared to misidentified HJs.

3. DIFFERENCES BETWEEN THE DOPPLER AND KEPLER SAMPLES

While there are numerous works on the HJ occurrence rate for the Doppler sample and the *Kepler* sample, we focus on two works that have the most consistency between samples, i.e., Wright et al. (2012) and Howard et al. (2012). Howard et al. (2012) used the *Kepler* sample to measure the occurrence rate of HJs, f_{Kepler} . Interpolating their result, the occurrence rate is $0.60\% \pm 0.10\%$ for HJs with $5.0 R_{\oplus} \leq R_p \leq 22.0 R_{\oplus}$ and $P < 10$ days. Emulating the Howard et al. sample, but using the California Planet Survey (CPS) sample, Wright et al. (2012) found that the occurrence rate of HJs for the Doppler sample is $1.20\% \pm 0.38\%$. Although marginally significant, Wright et al. (2012) speculated that the difference can be accounted for by metallicity and evolutionary stage difference between the Doppler and *Kepler* samples.

3.1. Metallicity Difference

The occurrence rate of HJs is a strong function of stellar metallicity (Gonzalez 1997; Santos et al. 2001, 2004; Fischer & Valenti 2005; Johnson et al. 2010; Sousa et al. 2011; Wang & Fischer 2013). A slight difference of stellar metallicity between the Doppler and *Kepler* sample will result in a significant

difference in the HJ occurrence rate. The CPS survey targets stars in the solar neighborhood. The metallicity distribution of these nearby stars can be obtained from the Spectroscopic Properties Of Cool Stars (SPOCS) catalog (Valenti & Fischer 2005). We apply the effective temperature and surface gravity cut ($4100 \text{ K} \leq T_{\text{eff}} \leq 6100 \text{ K}$, $4.0 \leq \log g \leq 4.9$) to the SPOCS catalog to select solar-type stars. The cut is consistent with the solar-type star definition from Howard et al. (2012) for a proper comparison to the *Kepler* sample. A total of 694 stars from the SPOCS are selected. The mean and median metallicity for the Doppler sample is -0.01 and 0.02 dex. In comparison, the mean and median metallicity of the *Kepler* sample -0.04 and -0.03 dex for 12,400 *Kepler* stars (Dong et al. 2014). The metallicity difference between the Doppler sample and the *Kepler* sample is thus 0.03 dex (mean) or 0.05 dex (median), which results in a factor of 1.15 (mean) or 1.26 (median) difference in the HJ occurrence rate if using a power law of 2.0 (Fischer & Valenti 2005), since not only the mean/median but also the shape of the metallicity distribution can affect the HJ occurrence rate. For example, for two samples with the same mean/median but different widths of distribution, the wider sample will have a higher HJ occurrence rate because it contains a higher fraction of super metal-rich stars, which increases the HJ occurrence rate. After considering the shape of the metallicity distribution, the HJ occurrence rate of the Doppler sample should be 1.07 times higher than that of the *Kepler* sample. Therefore, the stellar metallicity alone cannot account for the difference in the HJ occurrence rate. Given the sample size for metallicity determination, i.e., 694 for the Doppler sample and 12,400 for the *Kepler* sample, the standard error in the metallicity measurement is small. However, the systematic error in the metallicity measurement is estimated at

~ 0.05 dex (Hinkel et al. 2014), which is much higher than the standard error. Taking the systematic error into account, stellar metallicity may lead to a factor of 1.35 difference, but still is not adequate to explain the factor of two to three difference in the HJ occurrence rate.

3.2. Evolutionary Stage Difference

HJs occur less frequently around evolved stars than around main-sequence stars (Johnson et al. 2010; Bowler et al. 2010). Inclusion of evolved stars may lower the HJ occurrence rate for a survey. The CPS uses SPOCS as the input catalog. Following the selection criteria of Wright et al. (2012), $V < 8$, $B - V < 1.2$, and $\Delta M_V < 2.5$ mag. The latter requirement corresponds roughly to $\log g > 3.5$. There are 745 stars in SPOCS meeting the criteria. Among these stars, 12.6% of stars have $\log g < 4.1$. The range $3.5 < \log g < 4.1$ is what is defined as subgiants in Bastien et al. (2014) for the *Kepler* sample. They found that 48% of bright *Kepler* stars are subgiants. Therefore, the fraction of subgiants is much higher for the *Kepler* sample than for the Doppler sample. The difference in the subgiant fraction can lead to the difference in the HJ occurrence rate.

3.3. Stellar Multiplicity Rate Difference

Doppler planet surveys usually target single stars and stars without nearby ($\text{sep} < 2''$) stellar companions (Wright et al. 2012); thus, the stellar multiplicity rate for the Doppler sample should be lower than what is known for stars in the solar neighborhood (Duquennoy & Mayor 1991; Raghavan et al. 2010). Given the median distance of 30 pc for the stars selected in Wright et al. (2012), a separation of $2''$ corresponds to 60 AU, which is roughly the peak of the stellar separation distribution (Raghavan et al. 2010). Therefore, the stellar multiplicity rate for the Doppler sample is at most half of the stellar multiplicity rate in the solar neighborhood, i.e., $\sim 23\%$. In comparison, stars in the *Kepler* sample are, in general, much further away at $\sim 200\text{--}1000$ pc. There is no strong selection bias against stars with stellar companions (Brown et al. 2011). The *Kepler* pixel scale is $4''$ and photometric aperture is usually at least twice as large; thus, the *Kepler* mission observes both single stars and multiple stellar systems with stellar separations that almost extend to the tail of stellar separation distribution. Therefore, the stellar multiplicity rate of the *Kepler* sample should be higher than the Doppler sample. Given that planet formation is suppressed in multiple star systems (e.g., Wang et al. 2014a), including more multiple star systems in the sample may lead to a lower planet occurrence rate.

3.4. False Positive Rate for HJs from Kepler

Not all HJ candidates detected by the *Kepler* are bona fide planets. The false positive rate (FPR) of HJs is estimated between 10% and 35% (Morton & Johnson 2011; Santerne et al. 2012; Fressin et al. 2013). Considering the FPR for HJs, the HJ occurrence rate from the *Kepler* mission should be even lower. Fressin et al. (2013) found the HJ occurrence rate to be 0.43% accounting for the FPR.

4. PROBING THE FUNDAMENTAL HJ OCCURRENCE RATE

We have covered a variety of potential causes that may account for the difference of HJ occurrence rates from the Doppler and *Kepler* sample. The following equations summarize how

Table 1
Current Knowledge of Parameters in Equations (4) and (5) and the Differences of Stellar Properties and Stellar Multiplicity between the Doppler Sample and the *Kepler* Sample

	<i>Kepler</i>	Doppler
f	✓	✓
$\xi_{[\text{Fe}/\text{H}]}$		✗
$\xi_{\text{misidentified}}$		✗
FPR	✗	✓
SGR	✗	✓
MR _{MS}	✗	✗
MR _{SG}	✗	✗

Notes. Parameters for which we have a good handle are marked with ✓. Parameters that are relatively uncertain are marked with ✗. These parameters are as follows. f : measured HJ occurrence rate. $\xi_{[\text{Fe}/\text{H}]}$: correcting factor for the metallicity difference. $\xi_{\text{misidentified}}$: correcting factor for misidentified HJs due to photometric dilution and subgiant contamination. FPR: false positive rate of *Kepler* planet candidates. SGR: fraction of subgiants. MR_{MS}: stellar multiplicity rate for main-sequence stars. MR_{SG}: stellar multiplicity rate for subgiants.

these potential causes can be combined to probe the fundamental HJ occurrence rate as a function of stellar multiplicity and the evolutionary stage. The measured HJ occurrence rate from the *Kepler* or the Doppler survey is a combined result of fundamental HJ occurrence rates:

$$f = f_{\text{MS}} \times (1 - \text{SGR}) + f_{\text{SG}} \times \text{SGR}, \quad (4)$$

$$f = f_{\text{single}} \times (1 - \text{MR}) + f_{\text{multiple}} \times \text{MR}, \quad (5)$$

where f is the measured HJ occurrence rate, MS represents the main sequence, SG represents the subgiant, MR is the stellar multiplicity rate, and SGR is the fraction of subgiants in the sample. We have two measurements from the *Kepler* and Doppler surveys. However, there are four fundamental HJ occurrence rates that we wish to solve for, $f_{\text{MS, single}}$, $f_{\text{MS, multiple}}$, $f_{\text{SG, single}}$, and $f_{\text{SG, multiple}}$. In addition, there are correcting factors for the difference between the *Kepler* and Doppler sample. For example, we should consider $\xi_{[\text{Fe}/\text{H}]}$, the correction factor for the metallicity difference (Section 3.1), $\xi_{\text{misidentified}}$, the correction factor for the misidentified HJs due to photometric dilution and subgiant contamination (Section 2.5), and FPR, the false positive rate for HJs from the *Kepler* sample (Section 3.4).

We wish to solve for these fundamental HJ occurrence rates. However, current knowledge of the *Kepler* and Doppler sample is inadequate. Table 1 summarizes our current knowledge of the parameters in the above equations. Different Doppler planet surveys converge to $f_{\text{Doppler}} \sim 1\%$ for the HJ occurrence rate (Mayor et al. 2011; Cumming et al. 2008; Wright et al. 2012). f_{Kepler} measured from different independent works agree at $\sim 0.4\%\text{--}0.6\%$ (Howard et al. 2012; Fressin et al. 2013). The correcting factor $\xi_{[\text{Fe}/\text{H}]}$ for the metallicity difference is still uncertain to ~ 0.05 dex which is the systematic error of the abundance analysis (Hinkel et al. 2014). The 0.05 dex difference would result in 25% change in the HJ occurrence rate. We quantify $\xi_{\text{misidentified}}$ in this work, but its value depends on assumptions of other parameters, such as MR and SGR. FPR for the *Kepler* sample has been studied extensively, but its value is still not well constrained, ranging from 10% to 35% (Morton & Johnson 2011; Santerne et al. 2012; Fressin et al. 2013).

Understanding the stellar properties and statistics is crucial in solving Equations (4) and (5). The determination of SGR requires the measurement of $\log g$. While the stars in the Doppler sample have relatively well-determined $\log g$ (Valenti & Fischer 2005; Sousa et al. 2011), $\log g$ distribution for the *Kepler* sample is still uncertain especially for the faint end (Bastien et al. 2014). MR_{MS} for the Doppler sample is not well constrained because of selection bias. While there is no strong selection bias against multiple star systems for the *Kepler* sample, but it is not known whether MR_{MS} is the same for the *Kepler* sample as for the solar neighborhood (Raghavan et al. 2010). The knowledge of MR_{SG} is more uncertain for both the *Kepler* and Doppler sample.

5. HJ OCCURRENCE RATE IN THE SOLAR NEIGHBORHOOD

It is very challenging, if not impossible, to probe the fundamental HJ occurrence rates given the uncertainties of the parameters required for the calculation. In addition, we need some additional constraints to solve for the four variables, $f_{MS,single}$, $f_{MS,multiple}$, $f_{SG,single}$, and $f_{SG,multiple}$, with only two measurements from the Doppler and *Kepler* sample. However, the fundamental HJ occurrence rates can be calculated under two extreme and yet unlikely circumstances where the four variables are reduced to two. While the following two extreme cases do not necessarily reflect the reality, but they help us to understand how we can use Equations (4) and (5) to probe the fundamental HJ occurrence rates.

5.1. Extreme Case I: $f_{single} = f_{multiple}$

It is still under debate how the HJ occurrence rate in single stars compares to that for multiple stellar systems. On one hand, evidence of suppressed planet formation is found in multiple stellar systems (e.g., Wang et al. 2014a, 2014b). On the other hand, a stellar companion may facilitate the formation of HJs via Kozai perturbation (e.g., Wu & Murray 2003), although a recent study shows that HJ formation via this channel may have an upper limit of 44% (Dawson et al. 2012). Therefore, it is not an unreasonable assumption that the HJ occurrence rate is the same in single stars as in multiple stellar systems.

With this assumption, Equations (4) and (5) are reduced to Equation (4) with two variables, f_{MS} and f_{SG} . Two measurements of f are available. One is from the *Kepler* sample and the other one is from the Doppler sample. In this calculation, we adopt $f_{Kepler} = 0.6\% \pm 0.1\%$ (Howard et al. 2012), $f_{Doppler} = 1.20\% \pm 0.38\%$ (Wright et al. 2012). We apply correction factors $\xi_{[Fe/H]} = 1.26$, $\xi_{misidentified} = 1.14$, and $FAP = 17\%$ (Fressin et al. 2013). We adopt SGR for the *Kepler* sample to be 48% (Bastien et al. 2014), and 12.6% for the Doppler sample (Valenti & Fischer 2005). Substituting all of the adopted values into Equation (4) for the *Kepler* and Doppler sample, we infer that $f_{MS} = 1.37\% \pm 0.52\%$ and $f_{SG} = 0.00\% \pm 0.61\%$. The uncertainties for f_{MS} and f_{SG} are based only on the uncertainties of f_{Kepler} and $f_{Doppler}$. The result indicates that HJs are much less common around subgiant stars than around main-sequence stars, which is consistent with observations (Bowler et al. 2010; Johnson et al. 2010).

5.2. Extreme Case II: $f_{MS} = f_{SG}$

On the other hand, Doppler observations of subgiants have not completely ruled out the possibility that HJs may be as common around subgiants as around main-sequence stars. The sample in Bowler et al. (2010) contains 31 subgiants. It is difficult to

rule out that $f_{MS} \neq f_{SG}$ with the small sample size. If we assume that $f_{MS} = f_{SG}$, Equations (4) and (5) are reduced to Equation (5) with two variables, f_{single} and $f_{multiple}$. In this case, we adopt $MR = 46\%$ for the *Kepler* sample, which is the same as the stellar multiplicity rate in the solar neighborhood (Raghavan et al. 2010). We adopt $MR = 5\%$, which is the stellar multiplicity rate for all the known planets detected by the Doppler technique (Wright et al. 2011). Adopting the same values for other parameters as the previous case, we infer that $f_{single} = 1.26\% \pm 0.43\%$ and $f_{multiple} = 0.08\% \pm 0.55\%$. The uncertainties for f_{single} and $f_{multiple}$ are based only on the uncertainties of f_{Kepler} and $f_{Doppler}$.

We emphasize that the two extreme cases do not necessarily reflect the reality. In fact, one extreme assumption leads to a result that is significantly different from the other extreme assumption. Therefore, a meaningful solution for these fundamental HJ occurrence rates must lie in between these two extremes. To better understand the fundamental HJ occurrence rate, we need to have better knowledge of the stellar properties of the *Kepler* and Doppler samples and some other external constraints.

6. SUMMARY AND DISCUSSION

6.1. Summary

We conduct a simulation to investigate the fraction of HJs missed or misidentified by the *Kepler* mission due to the effects of photometric dilution and subgiant contamination. Despite these two effects, the *Kepler* mission rarely missed any HJs because of their large photometric signal. However, $12.48\% \pm 0.24\%$ of HJs are misidentified as smaller planets due to photometric dilution and subgiant contamination. Therefore, the occurrence rate of HJs from the *Kepler* mission should be revised upward by $\sim 14\%$.

We review the differences of stellar properties and stellar multiplicity between the *Kepler* and Doppler sample. The differences, together with the misidentified HJs, can potentially explain the discrepancy in the HJ occurrence rate. We discuss the fundamental HJ occurrence rate as a function of stellar multiplicity and evolutionary stage, i.e., HJ occurrence rate for single and multiple stellar systems and for main-sequence and subgiant stars. We attempt to solve for the fundamental HJ occurrence rate by synthesizing the results from the *Kepler* and Doppler sample. However, our current knowledge of stellar properties and stellar multiplicity rate is too limited for us to draw any quantitative conclusion. In two special cases, though not entirely realistic, we find qualitative evidence that HJs occur less frequently around subgiant stars and in multiple stellar systems.

6.2. Future Work on the Fundamental HJ Occurrence Rates

To quantitatively calculate the fundamental HJ occurrence rate, $f_{MS,single}$, $f_{MS,multiple}$, $f_{SG,single}$, and $f_{SG,multiple}$, we must need external constraints in addition to the Doppler and *Kepler* results. For example, one can calculate $f_{MS,single}$ by carefully selecting main-sequence single stars in the Doppler survey. Alternatively, one can select main-sequence stars and subgiants in the *Kepler* sample based on the “flicker” method (Bastien et al. 2014) or spectroscopic methods. From these two sample of *Kepler* stars, one can calculate the ratio of f_{MS} and f_{SG} assuming similar stellar multiplicity rates for main-sequence and subgiant stars. In addition, the ratio of $f_{MS,single}$ to $f_{MS,multiple}$ is provided by Wang et al. (2014a), although the caveat is that the ratio is for a sample of planet candidates that are mostly small planets. With

more constraints, the fundamental planet occurrence may be calculated quantitatively. However, much more effort is needed to make sure the calculation is meaningful. Future missions such as K2 (Howell et al. 2014) and TESS (Ricker et al. 2014) may provide independent measurement of HJ occurrence rate, i.e., f_{K2} and f_{TESS} , these results can be incorporated to calculate the fundamental HJ occurrence rate, given that stellar properties and the stellar multiplicity rate are known to a certain accuracy.

6.3. Implications to η_{Earth}

The findings in this paper can be used for the determination of η_{Earth} , the occurrence rate of an Earth-like planet in the habitable zone. This issue has already been complicated by the definition of the habitable zone (Seager 2013; Kopparapu et al. 2013). Furthermore, the determination of η_{Earth} should take into account the photometric dilution and subgiant contamination. Finally, the measurement of η_{Earth} from either the *Kepler* or the Doppler survey is a combination of the fundamental η_{Earth} as a function of stellar multiplicity rate and evolutionary stage (Equations (4) and (5)). These fundamental η_{Earth} can be precisely determined only if we have a good understanding of the stellar properties and stellar multiplicity rate of the sample.

We thank the referee Ronald Gilliland for his insightful comments and suggestions that substantially improved the paper. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

REFERENCES

- Bastien, F. A., Stassun, K. G., & Pepper, J. 2014, *ApJL*, 788, L9
 Bayliss, D. D. R., & Sackett, P. D. 2011, *ApJ*, 743, 103
 Bonfils, X., Delfosse, X., Udry, S., et al. 2013, *A&A*, 549, A109
 Bowler, B. P., Johnson, J. A., Marcy, G. W., et al. 2010, *ApJ*, 709, 396
 Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, *AJ*, 142, 112
 Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, *PASP*, 120, 531
 Dawson, R. I., Murray-Clay, R. A., & Johnson, J. A. 2012, arXiv:1211.0544
 Dong, S., Zheng, Z., Zhu, Z., et al. 2014, *ApJL*, 789, L3
 Duquennoy, A., & Mayor, M. 1991, *A&A*, 248, 485
 Feiden, G. A., & Chaboyer, B. 2012, *ApJ*, 757, 42
 Fischer, D. A., & Valenti, J. 2005, *ApJ*, 622, 1102
 Fressin, F., Torres, G., Charbonneau, D., et al. 2013, *ApJ*, 766, 81
 Girardi, L., Groenewegen, M. A. T., Hatziminaoglou, E., & da Costa, L. 2005, *A&A*, 436, 895
 Gonzalez, G. 1997, *MNRAS*, 285, 403
 Gould, A., Dorsher, S., Gaudi, B. S., & Udalski, A. 2006, *AcA*, 56, 1
 Henry, T. J., Ianna, P. A., Kirkpatrick, J. D., & Jahreiss, H. 1997, *AJ*, 114, 388
 Hinkel, N. R., Timmes, F. X., Young, P. A., Pagano, M. D., & Turnbull, M. C. 2014, *AJ*, 148, 54
 Horch, E. P., Howell, S. B., Everett, M. E., & Ciardi, D. R. 2014, *ApJ*, 795, 60
 Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, *ApJS*, 201, 15
 Howell, S. B., Sobek, C., Haas, M., et al. 2014, *PASP*, 126, 398
 Huber, D., Silva Aguirre, V., Matthews, J. M., et al. 2014, *ApJS*, 211, 2
 Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010, *Proc. SPIE*, 7740, 77400D
 Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010, *PASP*, 122, 905
 Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, *ApJ*, 765, 131
 Kraus, A. L., & Hillenbrand, L. A. 2007, *AJ*, 134, 2340
 Lissauer, J. J. 1993, *ARA&A*, 31, 129
 Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., et al. 2011, *ApJS*, 197, 8
 Marcy, G., Butler, R. P., Fischer, D., et al. 2005, *PTHPs*, 158, 24
 Marcy, G. W., & Butler, R. P. 1996, *ApJL*, 464, L147
 Mayor, M., & Queloz, D. 1995, *Natur*, 378, 355
 Mayor, M., Marmier, M., Lovis, C., et al. 2011, arXiv:1109.2497
 Mortier, A., Santos, N. C., Sozzetti, A., et al. 2012, *A&A*, 543, A45
 Morton, T. D., & Johnson, J. A. 2011, *ApJ*, 738, 170
 Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, *ApJS*, 190, 1
 Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, *Proc. SPIE*, 9143, 20
 Santerne, A., Díaz, R. F., Moutou, C., et al. 2012, *A&A*, 545, A76
 Santos, N. C., Israelian, G., & Mayor, M. 2001, *A&A*, 373, 1019
 Santos, N. C., Israelian, G., & Mayor, M. 2004, *A&A*, 415, 1153
 Schlaufman, K. C., & Winn, J. N. 2013, *ApJ*, 772, 143
 Seager, S. 2013, *Sci*, 340, 577
 Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M., & Udry, S. 2011, *A&A*, 533, A141
 Udry, S., & Santos, N. C. 2007, *ARA&A*, 45, 397
 Valenti, J. A., & Fischer, D. A. 2005, *ApJS*, 159, 141
 Wang, J., & Fischer, D. A. 2013, *AJ*, 149, 14
 Wang, J., Fischer, D. A., Xie, J.-W., & Ciardi, D. R. 2014a, *ApJ*, 791, 111
 Wang, J., Xie, J.-W., Barclay, T., & Fischer, D. A. 2014b, *ApJ*, 783, 4
 Wright, J. T., Fakhouri, O., Marcy, G. W., et al. 2011, *PASP*, 123, 412
 Wright, J. T., Marcy, G. W., Howard, A. W., et al. 2012, *ApJ*, 753, 160
 Wu, Y., & Murray, N. 2003, *ApJ*, 589, 605