Draft version February 23, 2018
Typeset using LaTeX **modern** style in AASTeX61

BINARITY'S BIASES ON TRANSIT SURVEY OCCURRENCE RATES

L. G. Bouma,[1] K. Masuda,[1] and J. N. Winn[1]

[1] *Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540, USA*

Submitted to AAS journals.

ABSTRACT

Wide-field surveys for transiting planets, such as the NASA *Kepler* and *TESS* missions, are usually conducted without knowing which stars have binary companions. Unresolved and unrecognized binaries give rise to systematic errors in planet occurrence rates, including misclassified planets and mistakes in completeness corrections. The individual errors can have different signs, making it difficult to anticipate the net effect on inferred occurrence rates. Here we use simplified models of signal-to-noise limited transit surveys to try and clarify the situation. We derive a formula for the apparent occurrence rate density measured by an observer who falsely assumes all stars are single. The formula depends on the binary fraction; the mass function of the secondary stars; and the true occurrence of planets around primaries, secondaries, and single stars. It also takes into account the Malmquist bias by which binaries are over-represented in flux-limited samples. Application of the formula to an idealized *Kepler*-like survey shows that for planets larger than $2\,R_\oplus$, the net systematic error is of order 5%. In particular, unrecognized binaries are unlikely to be the reason for the apparent discrepancies between hot Jupiter occurrence rates measured in different surveys. For smaller planets the size the errors are potentially larger: the occurrence of Earth-sized planets could be overestimated by as much as 50%. We also show that whenever high-resolution imaging reveals a transit host star to be a binary, the planet is usually more likely to orbit the primary star than the secondary star.

Keywords: methods: data analysis — planets and satellites: detection — surveys

Corresponding author: L. Bouma
luke@astro.princeton.edu

1. INTRODUCTION

One of the goals of exoplanetary science is to establish how common, or rare, are planets of various types. Knowledge of planet occurrence rates is helpful for inspiring and testing theories of planet formation, designing the next generation of planet-finding surveys, and simply satisfying our curiosity. One method for measuring occurrence rates is to monitor the brightnesses of many stars over a wide field, seeking evidence for planetary transits. This was the highest priority of the NASA *Kepler* mission. Great strides have been made in the analysis of *Kepler* data, including progress towards measuring the fraction of Sun-like stars that harbor Earth-like planets (Youdin 2011; Petigura et al. 2013; Dong & Zhu 2013; Foreman-Mackey et al. 2014; Burke et al. 2015).

A lingering concern about these studies is that in most cases, investigators have assumed that all of the targets in the survey are single stars (*e.g.*, Howard et al. 2012; Fressin et al. 2013; Dressing & Charbonneau 2015; Burke et al. 2015). In reality, many of the sources that are monitored in a transit survey are unresolved multiple-star systems, mainly binaries. Unrecognized binaries cause numerous systematic errors in the planetary occurrence rates. For example, when there is a transiting planet around a star in a binary, the additive constant light from the second star reduces the fractional loss of light due to the planet. This makes transit signals harder to detect and lowers the number of detections. On the other hand, a binary system presents two opportunities to detect transiting planets, which could increase the overall number of detections.

At the outset of this study it was not clear to us whether the neglect of binaries is a serious problem, or even whether the net effect of the errors is positive or negative. The goals of this study were to provide a framework for dealing with these issues, and to gauge at least the order of magnitude of the systematic errors. In this spirit, our models are idealized. We do not attempt a detailed correction of the results from *Kepler* or any other real transit survey. We took inspiration from the insightful analytic model for transit surveys by Pepper et al. (2003).

This paper is organized as follows. The next section enumerates the various errors that arise from unrecognized binaries. Then in Section 3, we develop an idealized model of a transit survey in which all planets have identical properties, and all stars are identical except that some fraction are in binary systems. This simple model motivates the derivation of a general formula, given in Section 3.3, that allows for more realistic stellar and planetary populations. We use this formula in Section 4 to explore more realistic models. We discuss the errors due to unrecognized binaries for specific cases of current interest: the occurrence of Earth-like planets; the apparent discrepancy between hot Jupiter occurrence rates measured in different surveys; and the shape of the "evaporation valley" in the planet radius distribution that was brought to light by Fulton et al. (2017). We summarize and discuss all the results in Section 5.

2. UNDERSTANDING THE ERRORS

Imagine that a group of astronomers wants to measure the mean number of planets per star. They are particularly interested in planets of radius r and stars of mass M and radius R. They obtain a time series of images of some region of the sky, and prepare light curves for a large number of unresolved sources. Then they search these light curves for transit signals and detect all the signals for which

$$\frac{\delta}{\sigma} > \left(\frac{\mathrm{S}}{\mathrm{N}}\right)_{\mathrm{min}}. \tag{1}$$

Here the signal, δ, is the observed transit depth, the dimensionless fraction by which the total light fades during transits. Note that although δ is often equated with $(r/R)^2$, this is not true when the host star is a member of an unresolved binary. In those cases, δ is smaller than $(r/R)^2$ because of the constant light from the binary companion, an effect often called the "dilution" of the transit signal. The noise, σ, is the fractional uncertainty in the determination of the flux of the source, which may include multiple stars that are blended together. The threshold signal-to-noise ratio depends on the desired level of confidence that the signal is real.

The astronomers analyze their data assuming that all the sources are single stars. In particular they do not have accurate enough parallaxes to tell that some of the stars appear to be overluminous. They count the number N_{det} of transit signals that appear to be produced by the desired type of planet around the desired type of star. They also count the number $N_\star$ of "searchable stars" in their survey, i.e., the number of stars of the desired type that are bright enough to have allowed for the detection of a transit signal with amplitude $(r/R)^2$. They estimate the occurrence rate to be

$$\Lambda = \frac{N_{\mathrm{det}}}{N_\star}\frac{1}{p_{\mathrm{tra}}}, \tag{2}$$

where the geometric transit probability, p_{tra}, accounts for the fact that most planetary orbits are not aligned close enough with our line of sight to produce transits.

There are many potential pitfalls in this calculation. Some genuine transit signals are missed even if they formally exceed the signal-to-noise threshold, because of the probabilistic nature of transit detection. Planets can be misclassified due to statistical and systematic errors in the catalogued properties of the stars. Some transit-like signals are spurious, arising from noise fluctuations or failures of "detrending" the astrophysical or instrumental variations in the photometric signal. Poor angular resolution leads to blends between eclipsing binary stars and other stars along nearly the same line of sight, producing signals that mimic those of transiting planets.

Here, though, we will focus exclusively on problems that arise from the fact that many stars exist in gravitationally bound binary systems. We will also focus on the errors in planet occurrence as a function of radius, rather than orbital period. This is because when more than one transit is detected (as is usually required by the surveyors), the orbital periods can be measured without ambiguity regardless of whether the host star is single or one member of a binary. Even with this narrow focus, there are numerous sources of error. All three of the quantities in Equation 2 are biased:

1. The number of detected planets, $N_{\rm det}$, is actually the number of detected planets that *appear* to have size r, orbiting stars that *appear* to have mass M. Whenever the planet-hosting star is part of a binary,
   - the planet's size could be misclassified because of the reduction in the amplitude of the photometric signal;
   - the host star's properties could be misclassified because its light is combined with a second star of a different spectral type.
2. The number of searchable stars, $N_\star$, is biased
   - toward lower values, because it does not include all of the secondary stars that were inadvertently searched for transiting planets;
   - toward higher values, because some of the stars that appeared to be searchable are in fact binaries for which the amplitude of the photometric signal would have been reduced to an undetectable level.
3. The transit probability $p_{\rm tra}$ is biased because the planet-hosting star could be misclassified. At fixed orbital period, the transit probability is proportional to $\rho^{-1/3}$, where ρ is the stellar mean density (Winn 2010). Therefore, any errors in determining the host star's mean density lead to errors in the correction for the transit probability.

There are at least two other complications that may arise, which are not represented in Equation 2. The first one is an observational effect. Within the sample of apparently searchable stars, the ratio between the number of binary and single stars will differ from the ratio that would be found in a volume-limited sample. This is due to a type of Malmquist bias. The total luminosity of a binary is larger than the luminosity of either the primary or secondary star. This means that for transit signals of a given amplitude, sources that are binaries appear to be searchable at greater distances from the Earth. Binaries are therefore over-represented in the collection of apparently searchable stars.

The other complication is astrophysical: the true occurrence rate of a certain type of planet may depend on whether the host is a single star, the primary star of a binary, or the secondary star of a binary. The rate might also depend on the characteristics of the binary, such as the mass ratio and orbital period. Such differences could be caused by the requirement for long-term dynamical stability, or differences in the planet formation process. When the search sample includes both singles and binaries, the detected planets are thereby drawn from different occurrence distributions (see Wang et al. 2015a; Kraus et al. 2016).

Given all of the confusing and opposing sources of error, we will proceed in stages. We start with a model so simple that everything can be written down on the back of a napkin, and build up to an analytic model allowing for generality in the distribution of the binaries and the planets they host.

3. SIMPLE MODELS

3.1. *One type of star, one type of planet*

Since the effects of binarity are most pronounced when the two stellar components are similar, we begin by considering a universe in which all stars are identical, with mass M, radius R, and luminosity L. Single stars are uniformly distributed in space with a number density of $n_{\rm s}$ stars per cubic parsec, and binaries are uniformly distributed with number density $n_{\rm b}$. In this scenario, stars are never misclassified because the combined light of a binary has the same color and spectrum as a single star. We further assume that all planets have the same radius, r, and occur around single stars and members of binaries at the same rate, $\Lambda(r)$.

Our naive observers conduct a transit survey. To calculate the occurrence rate of planets with radius r, they count the number of detections of signals with amplitude $(r/R)^2$. Then they identify all the sources that appear to have been searchable for a signal of amplitude $(r/R)^2$. Here and throughout the rest of this paper, we assume that the limiting source of noise is the photon-counting noise from the source, i.e., $\sigma \propto 1/\sqrt{F}$, where where F is the total flux of the source. Thus the observers determine the minimum F_0 for which detection would have been possible, and count the number of sources with $F_{\rm tot} > F_0$. This will include all the single stars out to a maximum distance

$$d_0 = \sqrt{\frac{L}{4\pi F_0}}. \tag{3}$$

Since binaries are twice as luminous, the condition $F > F_0$ will include binaries out to the larger distance of $d_0\sqrt{2}$. None of the stars in binaries will appear to have a planet of radius r, because of the dilution of the transit signal. Thus, the *apparent* occurrence rate $\Lambda_{\rm a}$ of planets of radius r is

$$\Lambda_{\rm a}(r) = \frac{\Lambda(r)\, n_{\rm s} d_0^3}{n_{\rm s} d_0^3 + n_{\rm b}(d_0\sqrt{2})^3}. \tag{4}$$

This apparent rate is larger than the true rate by a factor

$$\frac{\Lambda_{\rm a}(r)}{\Lambda(r)} = \frac{1}{1 + 2^{3/2}(n_{\rm b}/n_{\rm s})}. \tag{5}$$

The observers will also detect some transiting planets around stars with binary companions. The amplitude of these signals is

$$\frac{L(r/R)^2}{L+L} = \frac{1}{2}\left(\frac{r}{R}\right)^2 = \left(\frac{r/\sqrt{2}}{R}\right)^2, \tag{6}$$

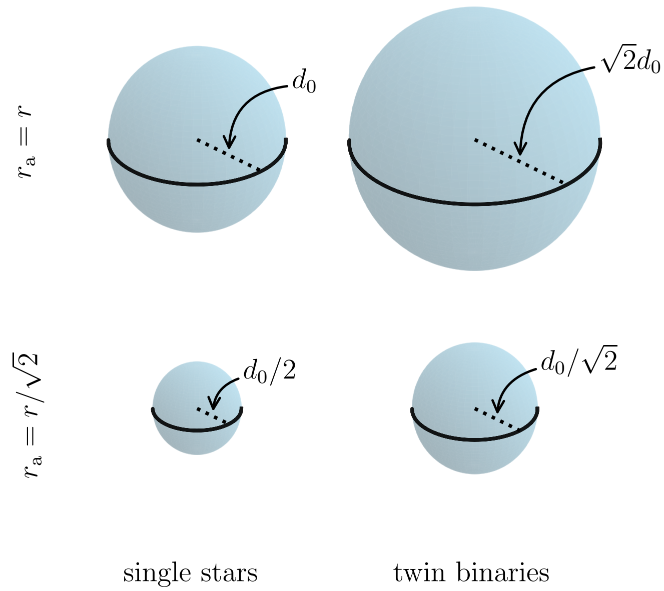

Figure 1. *Top.*—Volumes within which sources appear to be searchable for planets of radius r. Single stars are searchable out to a distance d_0, at which point they become too faint to allow the detection of transits. Binaries are brighter and therefore appear to be searchable out to a larger distance. *Bottom.*—Volumes within which stars appear to be searchable for planets of radius $r/\sqrt{2}$.

leading the astronomers to believe they have discovered a population of planets with radius $r/\sqrt{2}$. To calculate the corresponding occurrence rate, they count the stars for which this type of signal would have been detectable. The limiting flux for detection in this case is $4F_0$, because the signal amplitude hase been reduced by a factor of two and the noise level must also be reduced by a factor of two. The condition $F > 4F_0$ is met for single stars within a distance $d_0/2$, and binaries within a distance $d_0\sqrt{2}/2$. Therefore, the observers will calculate the occurrence rate of this new type of planet to be

$$\Lambda_a\left(\frac{r}{\sqrt{2}}\right) = \frac{2\Lambda(r)\, n_b(d_0\sqrt{2}/2)^3}{n_s(d_0/2)^3 + n_b(d_0\sqrt{2}/2)^3} = \frac{2\Lambda(r)\cdot 2^{3/2}(n_b/n_s)}{1 + 2^{3/2}(n_b/n_s)}. \tag{7}$$

Figure 1 illustrates the volumes enclosing the apparently searchable single and binary stars.

We can now assess the severity of the errors, for a given value of the binary-to-single ratio n_b/n_s. For stars with masses from 0.7 to 1.3 $M_\odot$, Raghavan et al. (2010) found the multiplicity fraction – the fraction of systems in a volume-limited sample that are multiple – to be 0.44. Assuming all multiple systems are binaries, this gives a binary fraction

$$\frac{n_b}{n_s + n_b} \approx 0.44, \tag{8}$$

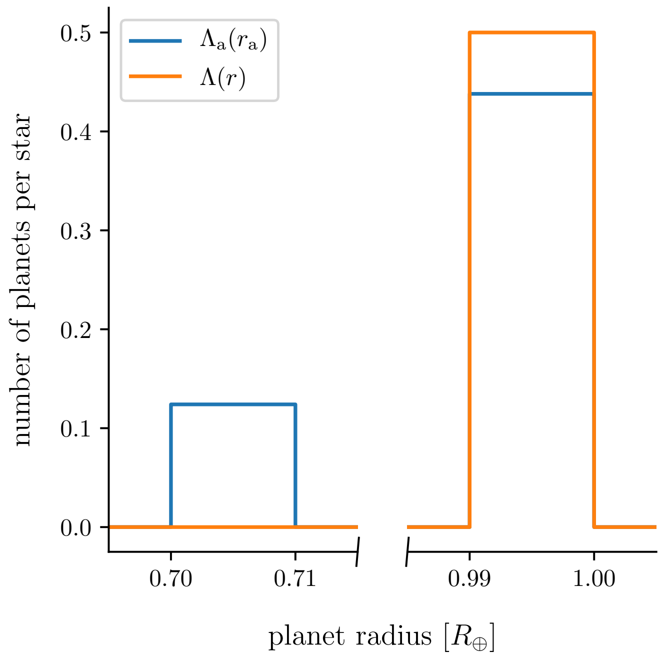

Figure 2. Apparent occurrence rate Λ_a, and true occurrence rate Λ in a universe with only one type of star, one type of planet (with radius $R_\oplus$), and a number density of binary stars 5% the number density of single stars. The occurrence rate of planets with radius $R_\oplus$ is underestimated, and the occurrence rate of planets with radius $R_\oplus/\sqrt{2}$ is overestimated.

which implies $n_b/n_s \approx 0.79$. Of course not all of these binaries are "twin" binaries as we have assumed in our simple calculation. Later, in Section 3.3, we will allow for a continuum of properties for the secondary stars. For now, we might guess that only a tenth of the binaries have pairs of stars close enough in brightness to produce errors as significant as those we have been considering. Adopting the value $n_b/n_s \approx 0.05$, we find

$$\frac{\Lambda_a(r)}{\Lambda(r)} = 0.88, \quad \frac{\Lambda_a(r/\sqrt{2})}{\Lambda(r)} = 0.25. \tag{9}$$

All together, the various effects produce biases of order 10% in the occurrence rates. We will see that this level of error is characteristic of many of our more complicated models as well.

3.2. *Planets of different sizes*

We will now generalize to allow for a continuum of planet sizes. We introduce the occurrence rate density

$$\Gamma(r) \equiv \frac{\mathrm{d}\Lambda}{\mathrm{d}r}, \tag{10}$$

the number of planets per star per unit planet radius. The naive observers are measuring $\Gamma_a(r_a)$, the apparent occurrence rate density of planets with an "apparent radius"

$$r_a = R\sqrt{\delta}. \tag{11}$$

For a given value of r_a, the observers assemble the sample of $N_\star(r_a)$ sources for which a signal of amplitude $\delta = (r_a/R)^2$ could have been detected. We showed in the previous section that when all stars are identical, this sample will contain binaries and single stars in the ratio $2^{3/2}(n_b/n_s)$. Here we will denote this ratio by μ, which will be calculated later in a more general context. With this definition, the number of single stars is proportional to $1/(1+\mu)$ and the number of binaries is proportional to $\mu/(1+\mu)$.

From within this sample of apparently searchable sources, the observers sum the number of detections $dN_{\rm det}$ of planets with an apparent radius between r_a and $r_a + dr_a$. There will be three contributions to this sum,

$$dN_{\rm det,0} = \frac{N_\star(r_a)p_{\rm tra}}{1+\mu}\Gamma_0(r_a){\rm d}r_a, \tag{12}$$

$$dN_{\rm det,1} = \frac{\mu N_\star(r_a)p_{\rm tra}}{1+\mu}\Gamma_1(\sqrt{2}r_a){\rm d}(\sqrt{2}r_a), \tag{13}$$

$$dN_{\rm det,2} = \frac{\mu N_\star(r_a)p_{\rm tra}}{1+\mu}\Gamma_2(\sqrt{2}r_a){\rm d}(\sqrt{2}r_a). \tag{14}$$

where here and elsewhere, the subscript 0 refers to single stars, 1 refers to primary stars of binaries, and 2 refers to secondary stars of binaries. The apparent occurrence rate density is therefore

$$\Gamma_a(r_a) = \frac{dN_{\rm det}}{dr_a}\frac{1}{N_\star(r_a)p_{\rm tra}} = \frac{\Gamma_0(r_a)}{1+\mu} + \frac{\mu\sqrt{2}\,\Gamma_1(\sqrt{2}r_a)}{1+\mu} + \frac{\mu\sqrt{2}\,\Gamma_2(\sqrt{2}r_a)}{1+\mu}. \tag{15}$$

Our earlier results, Equations 4 and 7, are recovered by integrating this formula over r after inserting

$$\Gamma_i(r) = \Lambda(r_p)\,\hat{\delta}(r - r_p), \quad \text{for } i \in \{0,1,2\} \tag{16}$$

where $\hat{\delta}$ is the Dirac delta function, and r_p is the single planet size considered in Section 3.1.

3.3. *Binaries with different mass ratios*

Next we generalize to allow for a spectrum of different properties for the secondary stars. For simplicity we assume that the stars form a one-parameter family specified by the stellar mass M. This is approximately the case for main-sequence stars. The functions $L(M)$ and $R(M)$ give the luminosity and radius as a function of mass, and $f(q)$ is the distribution of binary mass ratios in a volume-limited sample. We also assume that observers perceive all the binaries to be isolated stars with the same mass as the primary star, i.e., the light from the secondary star is either too faint or too similar to the primary star to make a difference in the stellar classification. Finally, for clarity of presentation we make the simplifying assumption that Γ_0, Γ_1 and Γ_2 do not depend on stellar mass, although our formalism can easily accommodate such a dependence.

To compute the apparent occurrence rate density, we need to make the following modifications to Equation 15:

1. The Malmquist bias is different. For a given value of $r_{\rm a}$, the number of binary systems in the searchable sample with mass ratio $(q, q+{\rm d}q)$ is

$$\frac{N_\star(r_{\rm a})}{1+\mu} \cdot \frac{n_{\rm b}}{n_{\rm s}} \left[\frac{L(M)+L(qM)}{L(M)}\right]^{3/2} f(q)\,{\rm d}q, \tag{17}$$

where μ is given by

$$\mu = \int_0^1 \frac{n_{\rm b}}{n_{\rm s}} \left[\frac{L(M)+L(qM)}{L(M_{\rm a})}\right]^{3/2} f(q)\,{\rm d}q. \tag{18}$$

To obtain the total number of searchable binaries, Equation 17 needs to be integrated over q. The result of the integration replaces $\mu N_\star(r_{\rm a})/(1+\mu)$ in Equations 13 and 14.

2. The apparent radius of a planet in a binary system now depends on whether the host star is the primary or the secondary star. When the host is the primary star, we write $r = \mathcal{D}_1 r_{\rm a}$, where

$$\mathcal{D}_1 = \left[\frac{L(M)+L(qM)}{L(M)}\right]^{1/2} \tag{19}$$

is the appropriate dilution factor. When the host is the secondary star, a correction must also be made to account for the different radius of the secondary star. In that case $r = \mathcal{D}_2 r_{\rm a}$, where

$$\mathcal{D}_2 = \frac{R(qM)}{R(M)} \left[\frac{L(M)+L(qM)}{L(qM)}\right]^{1/2}. \tag{20}$$

3. When a transiting planet is detected around a secondary star, the naive observers make the wrong correction for the transit probability. At a fixed orbital period, $p_{\rm tra}$ in Equation 14 must be multiplied by a factor of

$$\frac{R(qM)}{R(M)} q^{-1/3}. \tag{21}$$

Taking these modifications into account, a general formula for the apparent rate density is

$$\Gamma_{\rm a}(r_{\rm a}) = \frac{1}{1+\mu} \left\{ \Gamma_0(r_{\rm a}) + \frac{n_{\rm b}}{n_{\rm s}} \left[\int_0^1 {\rm d}q\, \mathcal{D}_1^3 f(q) \cdot \mathcal{D}_1 \Gamma_1(\mathcal{D}_1 r_{\rm a}) + \int_0^1 {\rm d}q\, \mathcal{D}_1^3 f(q) \cdot \mathcal{D}_2 \Gamma_2(\mathcal{D}_2 r_{\rm a}) \cdot \frac{R(qM)}{R(M)} q^{-1/3} \right] \right\}. \tag{22}$$

The appendix gives an alternative derivation of this equation.

We note that the combination $\mathcal{D}_1^3 f(q)$ is the mass-ratio distribution for binaries contained within the searchable sample of sources. With this in mind we may rewrite Equation 22 as

$$\Gamma_{\rm a}(r_{\rm a}) = \frac{1}{1+\mu}\left[\Gamma_0(r_{\rm a}) + \mu\left\langle \mathcal{D}_1 \cdot \Gamma_1\left(\mathcal{D}_1 r_{\rm a}\right) + q\mathcal{D}_2 \cdot \Gamma_2\left(\mathcal{D}_2 r_{\rm a}\right) \cdot \frac{R(qM)}{R(M)} q^{-1/3} \right\rangle\right], \tag{23}$$

where the angle brackets denote averaging over all the binaries in the searchable sample.

4. CASE STUDIES

We can now gauge the size of the systematic errors associated with unresolved binaries, for cases of interest. Throughout this section we assume $R \propto M$, and $L \propto M^\alpha$ where $\alpha = 3.5$, as is roughly the case for main-sequence stars. With these assumptions, Equations 19 and 20 become

$$\mathcal{D}_1 = (1+q^\alpha)^{1/2} \quad \text{and} \quad \mathcal{D}_2 = q(1+q^{-\alpha})^{1/2}. \tag{24}$$

4.1. *Power-law planet radius distribution*

Based on *Kepler* data, Howard et al. (2012) found the radius distribution of planets between 2 and 17 $R_\oplus$ orbiting Sun-like stars within 0.25 AU to be consistent with a power law,

$$\Gamma_0(r) \propto r^\gamma, \tag{25}$$

with $\gamma = -2.92 \pm 0.11$. Their analysis ignored the effects of binarity. We can use our formalism to estimate the resulting level of systematic error.

To warm up we will again consider the case in which all stars are identical. We assume further that the distributions of planets around primaries and secondaries differ only by multiplicative factors,

$$\Gamma_1 = Z_1\Gamma_0, \quad \Gamma_2 = Z_2\Gamma_0. \tag{26}$$

Application of Equation 22 gives

$$\frac{\Gamma_{\rm a}(r_{\rm a})}{\Gamma_0(r_{\rm a})} = \frac{1}{1+\mu} + 2^{\frac{\gamma+1}{2}}\frac{\mu}{1+\mu}\left(Z_1 + Z_2\right), \tag{27}$$

where $\mu = 2^{3/2} n_{\rm b}/n_{\rm s}$. The ratio does not depend on $r_{\rm a}$ and thus, under our assumptions, the effect of twin binaries is simply to change the normalization of the radius distribution. If we further assume that planet occurrence is independent of system multiplicity, i.e., $Z_1 = Z_2 = 1$, then

$$\frac{\Gamma_{\rm a}(r_{\rm a})}{\Gamma_0(r_{\rm a})} = \frac{1 + 2^{\frac{\gamma+3}{2}}\mu}{1+\mu}. \tag{28}$$

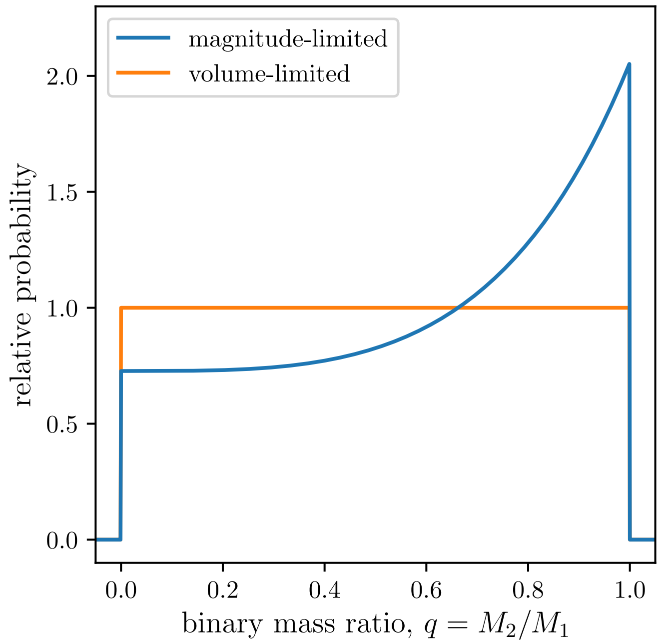

Figure 3. The mass ratio distribution of binaries in a magnitude-limited sample, assuming the underlying volume-limited distribution is uniform and the luminosity-mass relation is $L \propto M^{3.5}$. For transit signals of a given amplitude, the sample of searchable sources is magnitude-limited, causing high mass ratios to be over-represented.

Adopting $\gamma = -2.92$, and a twin binary fraction $n_\mathrm{b}/n_\mathrm{s} = 0.05$, as before, we obtain $\mu = 0.14$ and $\Gamma_\mathrm{a}/\Gamma_0 = 1.003$. The correction is tiny because, by coincidence, the reported value of γ is very close to -3, the value for which Equation 28 reduces to unity and the effects of binarity cancel out completely. This suggests that although Howard et al. (2012) derived the occurrence rate density under the false assumption that all stars are single, the resulting systematic error is negligible.

For a more accurate analysis we now consider a distribution of binary mass ratios. Studies of binaries in the local neighborhood suggest that the distribution of mass ratios is nearly uniform between zero and unity (Raghavan et al. 2010). We consider the more general possibility of a power-law dependence, $f(q) = \mathcal{N}_q q^\beta$, where $\mathcal{N}_q$ is a normalization constant. In this case Equation 18 gives

$$\mu = \frac{1}{\mathcal{N}_q}\frac{n_\mathrm{b}}{n_\mathrm{s}}\int_0^1 (1+q^\alpha)^{3/2} q^\beta \,\mathrm{d}q. \tag{29}$$

Figure 3 illustrates the Malmquist bias for the particular case of a flat distribution.

Again assuming $\Gamma_0 = \Gamma_1 = \Gamma_2$, the apparent occurrence rate density is

$$\frac{\Gamma_\mathrm{a}(r_\mathrm{a})}{\Gamma_0(r_\mathrm{a})} = \frac{1}{1+\mu}\left[1 + \frac{1}{\mathcal{N}_q}\frac{n_\mathrm{b}}{n_\mathrm{s}}\left(\int_0^1 \mathrm{d}q\, q^\beta (1+q^\alpha)^{\frac{\gamma+4}{2}} + \int_0^1 \mathrm{d}q\, q^{\beta+\gamma+\frac{5}{3}}(1+q^\alpha)^{\frac{3}{2}}(1+q^{-\alpha})^{\frac{\gamma+1}{2}}\right)\right], \tag{30}$$

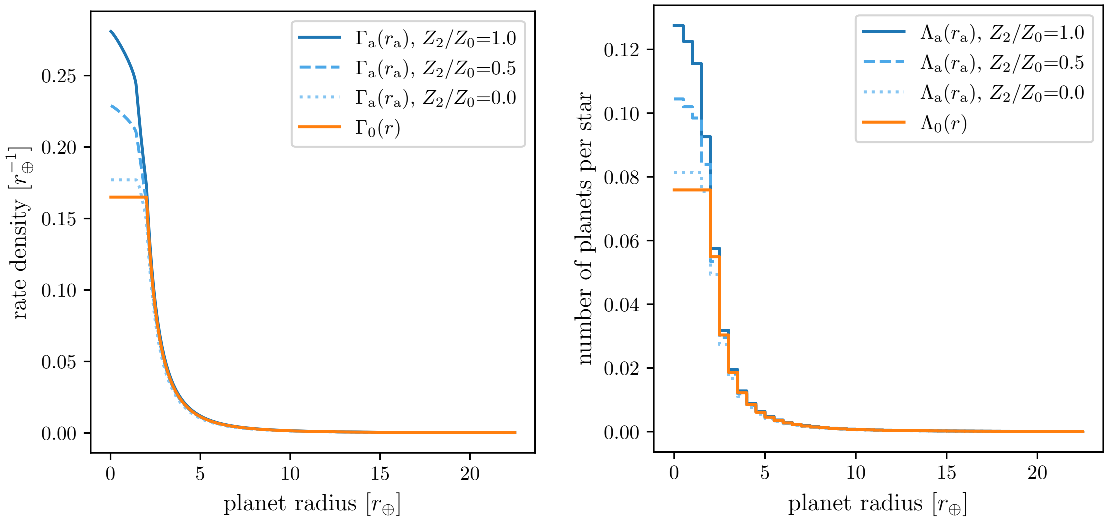

Figure 4. *Left:* Apparent occurrence rate density (Γ_a), compared to the true occurrence rate density (Γ_0) for single stars and primary stars in binaries. Three different cases are plotted, with different choices for the true occurrence rate density for secondary stars relative to that for single stars (Z_2/Z_0). *Right:* Same, but for apparent occurrence rate (Λ_a) within $0.5\,R_\oplus$ bins. In both cases the true planet radius distribution is specified by Equation 31.

Adopting the realistic numerical values $n_b/n_s = 0.79$, $\alpha = 3.5$, $\beta = 0$, and $\gamma = -2.92$, the summed integrals in Equation 30 give $(\ldots) \approx 1.503$, and $\Gamma_a/\Gamma_0 = 1.048$. The apparent occurrence rate density is 4.8% higher than the true value. We conclude that for planets around Sun-like stars with orbits within 0.25 AU and radii in the range 2–17 $R_\oplus$, the systematic errors associated with binarity are on the order of a few percent.

4.2. *Broken power-law planet radius distribution*

For planets smaller than $2\,R_\oplus$, the true occurrence rate density is more uncertain because such planets are more difficult to detect. To investigate the systematic errors associated with binarity on the inferred occurrence rate of Earth-sized planets, we need to make plausible assumptions about the true occurrence rate density. Obviously if the rate density continues to vary as r^{-3} down to much smaller planet sizes, the results of the preceding section will hold. But there is no particular reason to think this will be the case and indeed, some investigators have concluded that the rate density begins to level off to a constant value as r decreases below $R_\oplus$ (Petigura et al. 2013). To investigate the implications we consider a broken power-law:

$$f(r) \propto \begin{cases} r^{\gamma} & \text{for } r \geq 2R_\oplus \\ \text{constant} & \text{for } r \leq 2R_\oplus. \end{cases} \tag{31}$$

In this case the integrals that appear in Equation 22 are tedious to work out analytically, leading us to evaluate them numerically.[1] Figure 4 shows the results, again for the case $n_{\rm b}/n_{\rm s} = 0.79$, $\alpha = 3.5$, $\beta = 0$, and $\gamma = -2.92$.

The most obvious aspect of the results is that the occurrence rate density of planets smaller than the breakpoint radius of $2\,R_\oplus$ is now substantially overestimated. If secondary stars have the same planet population as the single stars, then the occurrence rate density of small planets is overestimated by 50%. The magnitude of the systematic error decreases if there are fewer planets around secondaries; if secondaries host no planets at all, the apparent rate is overestimated by only 10%.

The effects on larger planets are not as substantial. To be quantitative we consider the occurrence rate of giant planets with $r > 8\,R_\oplus$, by integrating the occurrence rate density displayed in Figure 4:

$$\Lambda_{\rm giant,a} = \int_{8\,R_\oplus}^{\infty} \Gamma_{\rm a}(r_{\rm a})\,{\rm d}r_{\rm a}, \quad \text{and} \quad \Lambda_{\rm giant,0} = \int_{8\,R_\oplus}^{\infty} \Gamma_0(r)\,{\rm d}r. \tag{32}$$

In this case the occurrence rate density is underestimated. The difference between the apparent and true rates is largest when the secondary stars do not host any planets, and has a magnitude $\Lambda_{\rm g,0}/\Lambda_{\rm g,a} = 1.13$. If instead the secondary stars have half as many giant planets as single stars, then the factor is reduced from 1.13 to 1.06.

4.3. *Which star has the planet?*

Our formalism also provides a way to calculate the relative probability that a detected planet orbits a single star, the primary star, or a secondary star. These can be obtained from the generalizations of Equations 12–14, which are derived in the appendix. This information can be used to help interpret the results of a transit survey. It is also relevant to the fairly common situation in which transits are detected from a source, and subsequent high-resolution imaging reveals the source to be a multiple-star system. The question then arises: which star is the host of the detected planet? Without further observations the answer is often unclear. To illustrate we return to the scenario described in the previous section: a broken power-law radius distribution with $\gamma = -2.92$, along with the choices $n_{\rm b}/n_{\rm s} = 0.79$, $\alpha = 3.5$, and $\beta = 0$. We also try different assumptions for the rate of planets around secondaries, i.e., different choices for Z_2/Z_0. We use Equations A18–A20 to compute the number of detections from single stars, primaries, and secondaries, and divide by the total number of detections at each apparent radius.

Figure 5 shows the results. When the apparent radius exceeds $2\,R_\oplus$, and planets exist about secondaries at the same rate as the primary star ($Z_2/Z_0 = 1$), then the planet is 3 times more likely to orbit the primary star than the secondary star. If secondaries host half as many planets as primaries ($Z_2/Z_0 = 0.5$), then a detected planet is 5 times more likely to orbit the primary star.

[1] We refer the interested reader to our online code for performing this computation: `github.com/lgbouma/binary_biases`, commit `6cb920c`.

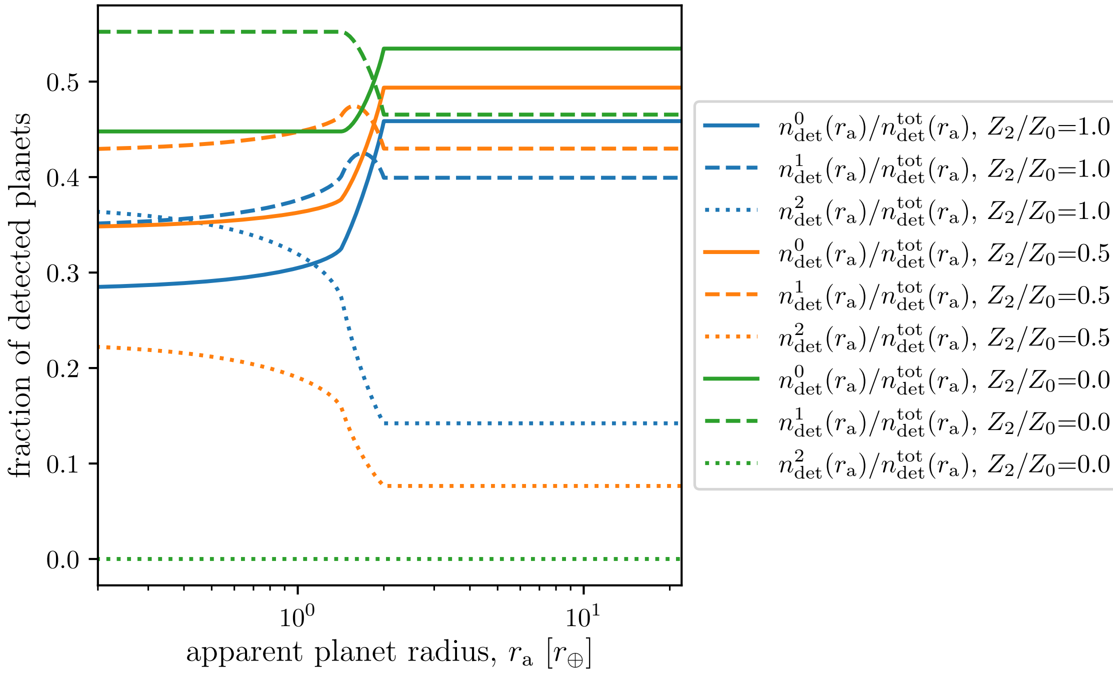

Figure 5. Fraction of detected planets that orbit single stars (solid lines), primary stars of binaries (dashed lines), and secondary stars of binaries (dotted lines), as a function of the apparent radius of the planet. We assume that single stars and primaries have the same number of planets per star. Three different cases are plotted, differing in the relative occurrence of planets around secondary stars (Z_2/Z_0). The true radius distribution is given by Equation 31.

The situation for smaller apparent radii is more nuanced. For $Z_2/Z_0 = 0$ or 0.5, any transit signals from binaries are still always more likely to arise from the primary star. However, if planets exist at the same rate in primaries and secondaries ($Z_2/Z_0 = 1$), then below apparent radii of $0.4\,R_\oplus$, more of the detected planets in binaries come from secondaries. They are actually much larger planets for which δ has been substantially reduced by the light from the primary star. As we consider apparent radii ranging from 2 to $1.4\,R_\oplus$, the fraction of detected planets with binary companions increases by anywhere from 6% to 12%, depending on the relative occurrence of planets about primaries and secondaries.

4.4. *A gap in the radius distribution*

Fulton et al. (2017) recently reported a "gap" in the radius distribution of close-in planets around Sun-like stars, between planet sizes of 1.5 and $2\,R_\oplus$. This can also be visualized as a "valley" in the occurrence rate density as a function of radius and period. The existence of the gap has been independently corroborated from a sample of KOIs with asteroseismically-determined stellar parameters (Van Eylen et al. 2017). Such a feature had been predicted as a consequence of the gradual photo-evaporation of the hydrogen-helium atmospheres of small rocky planets, during the first 100 Myr

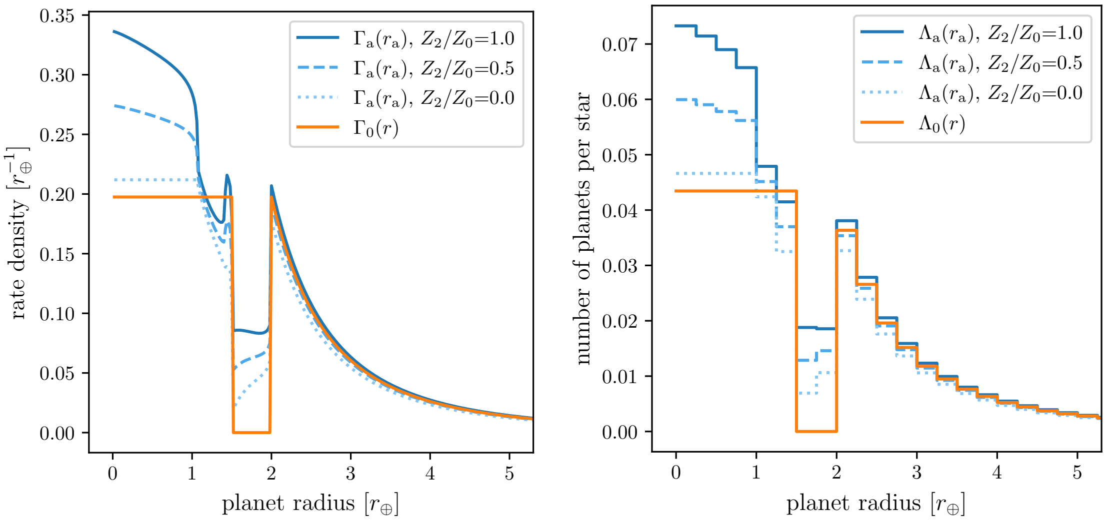

Figure 6. *Left:* True and apparent occurrence rate densities, and *right:* occurrence rates, as a function of planet radius, using a model for the radius distribution that exhibits a gap between 1.5 and 2 $R_\oplus$ (Equation 33).

of the age of the system when the host star produces a higher flux of ultraviolet and X-ray radiation (Owen & Wu 2017).

The occurrence calculations that led to this discovery, like almost all occurrence calculations, did not carefully account for unresolved binaries. Intuitively we expect the effect of unresolved binaries to have a blurring effect on the radius distribution, filling in any gaps in the radius distribution and making them appear less empty than in reality. To investigate the quantitative effects we make identical assumptions as in Section 4.2, except that the radius distribution is assumed to have a complete absence of planets with sizes between 1.5 and 2 $R_\oplus$:

$$f(r) \propto \begin{cases} r^\gamma & \text{for } r \geq 2\,R_\oplus, \\ 0 & \text{for } 1.5\,R_\oplus < r < 2\,R_\oplus, \\ \text{constant} & \text{for } r \leq 1.5\,R_\oplus. \end{cases} \tag{33}$$

Figure 6 shows the resulting true and apparent occurrence rate densities. For the case $Z_2/Z_0 = 1$, unresolved binaries reduces the contrast of the gap by nearly a factor of two, while also producing spurious features for apparent sizes beneath 1.5 $R_\oplus$. Thus the gap identified by Fulton et al. (2017) may be even more devoid of planets than it appears.

4.5. *Gaussian radius distribution*

The planet population may include some special members with a distinct radius distribution. For example in the recent study by Petigura et al. (2017b), hot Jupiters appear as an island in period-radius space, rather than as a component of a power-law distribution extending to smaller planets and longer periods. For this reason we test

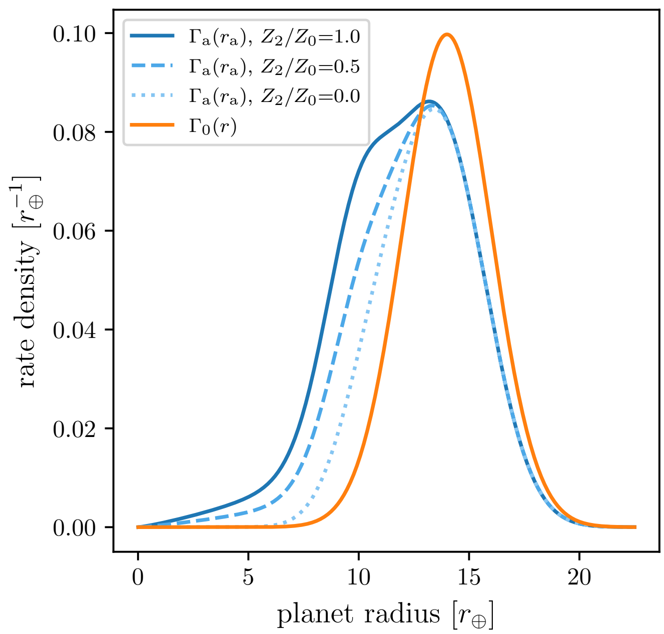

Figure 7. True and apparent occurrence rate densities, for a population of planets with true radii r drawn from a Gaussian distribution with mean $14\,R_\oplus$ and standard deviation $2\,R_\oplus$. This is similar to the hot Jupiter distribution presented by Petigura et al. (2017b).

the effects of unresolved binaries on a Gaussian radius distribution,

$$f(r) \propto \exp\left[-\frac{(r-\bar{r})^2}{2\sigma_r^2}\right], \tag{34}$$

with $\bar{r} = 14\,R_\oplus$ and $\sigma_r = 2\,R_\oplus$. As before, we allow for different normalizations of the planet populations around single, primary, and secondary stars: $\Gamma_i(r) = Z_i f(r)$. Figure 7 shows the results for the apparent radius distribution, $\Gamma_a(r_a)$. As one would expect, the effect is to smear the radius distribution toward smaller values, because some of the hot Jupiters now appear to be smaller planets. Less obviously, the integrated rate of hot Jupiters is also affected. We calculate the apparent hot Jupiter rate by integrating the apparent rate density for $r_a > 8\,R_\oplus$. If secondaries host any planets at all, the apparent rate is *greater* than the true rate. For instance, when hot Jupiters are just as common around secondary stars as single or primary stars, then $\Lambda_{\rm HJ,a}/\Lambda_{\rm HJ,0} = 1.23$. The apparent rate is higher because a larger number of stars were searched than are accounted for in the rate calculation.

5. DISCUSSION

How bad is ignoring binarity?—This study has shown that under a reasonable set of simplifying assumptions, ignoring binarity introduces systematic errors to star and planet counts in transit surveys, which then biases derived occurrence rates. Thus far, occurrence rate calculations[2] using transit survey data have mostly ignored stel-

[2] A list of occurrence rate papers is maintained at https://exoplanetarchive.ipac.caltech.edu/docs/occurrence_rate_papers.html

lar multiplicity (*e.g.*, Howard et al. 2012; Fressin et al. 2013; Foreman-Mackey et al. 2014; Dressing & Charbonneau 2015; Burke et al. 2015). For *Kepler* occurrence rates specifically, it seems that no one has yet carefully assessed binarity's importance, or lack thereof. This study does not resolve the problem; it only suggests the approximate scale of the systematic errors. Section 4.1 suggests that for apparent radii above $2r_\oplus$, binarity can be ignored down to a precision of a few percent. For apparent radii below $2r_\oplus$, the picture is less forgiving: Section 4.2 suggests that the apparent rates around single stars could be overestimated by as much as 50%.

The rate of Earth analogs—Youdin (2011), Petigura et al. (2013), Dong & Zhu (2013), Foreman-Mackey et al. (2014), and Burke et al. (2015) have found that between 0.03 and 1 Earth-like planets exist per Sun-like star, depending on assumptions that are made when calculating the rate (see Burke et al. 2015, Figure 17). Assuming a broken power-law, we showed that unrecognized binaries could cause overestimates in the rate of Earth analogs of up to 50%. This systematic effect is smaller than the other factors that currently dominate the dispersion in $\eta_\oplus$ measurements. If future analyses determine absolute values of $\eta_\oplus$ to better than a factor of two, binarity will likely merit closer attention.

One caveat is that none of our models included the rate density's period-dependence. However, binaries with separations $\lesssim 10\,\mathrm{AU}$ could provoke dynamical instabilities, leading to fewer Earth-like planets per star (*e.g.*, Holman & Wiegert 1999; Wang et al. 2014; Kraus et al. 2016). This would affect transit survey measurements of $\eta_\oplus$ beyond our rough estimate.

Hot Jupiter rate discrepancy—While unresolved binaries may bias $\eta_\oplus$ measurements, our approach suggests that they do not influence the hot Jupiter rate discrepancy. The "discrepancy" is that RV surveys find roughly twice as many hot Jupiters per Sun-like star as do transit surveys (roughly 1% vs. 0.5%, respectively). We summarize hot Jupiter rates reported by different surveys in Table 1.

Though the disagreement is only weakly significant ($< 3\sigma$), one reason to expect a difference is that the RV sample is biased towards metal-rich stars, which host more giant planets (Santos et al. 2004; Fischer & Valenti 2005; Gould et al. 2006). Investigating the discrepancy from the metallicty angle, Guo et al. (2017) measured the *Kepler* field's mean metallicity to be $[\mathrm{M/H}]_{\mathrm{Kepler}} = -0.045 \pm 0.009$, which is lower than the California Planet Search's mean of $[\mathrm{M/H}]_{\mathrm{CPS}} = -0.005 \pm 0.006$. The former value agrees with that found by Dong et al. (2014). Refitting for the metallicity exponent in $\Lambda_{\mathrm{HJ}} \propto 10^{\beta[\mathrm{M/H}]}$, Guo et al. found $\beta = 2.1 \pm 0.7$, and noted that the metallicity difference could account for 0.2% of the needed 0.5% difference between the measured CKS and *Kepler* rates[3]. Guo et al. concluded that "other factors, such as binary contamination and imperfect stellar properties" must also be at play.

[3] Petigura et al. (2017b) recently found $\beta = 3.4^{0.9}_{-0.8}$. If true, this implies that metallicity could account for about half of the hot Jupiter rate discrepancy.

Radial velocity surveys usually reject visual and spectroscopic binaries (Wright et al. 2012), so their hot Jupiter rates are closer to the rate for single stars than the rates reported by transit surveys. However, we found in Section 4.4 that when assuming a Gaussian radius distribution, apparent hot Jupiter rates are *greater* than the true rate around singles: the effect goes the wrong way. There could be other systematic factors behind the difference – but they are unlikely to be related to binarity.

Does a detected planet orbit the primary or secondary?—A separate reason to address stellar multiplicity in transit surveys is that it can radically alter the interpretation of planet candidates on a system-by-system level. To identify and correct for unseen stellar companions, high-resolution imaging campaigns have surveyed almost all *Kepler* Objects of Interest (Howell et al. 2011; Adams et al. 2012, 2013; Horch et al. 2012, 2014; Lillo-Box et al. 2012, 2014; Dressing et al. 2014; Law et al. 2014; Cartier et al. 2015; Everett et al. 2015; Gilliland et al. 2015; Wang et al. 2015b,c; Baranec et al. 2016; Ziegler et al. 2017). The results of these programs have been summarized by Furlan et al. (2017), and they represent an important advance in understanding the KOI sample's multiplicity statistics.

An immediate application is to reassess the radii of detected planets. Doing so, Hirsch et al. (2017) recently found that for their sample of KOIs in binaries, the planet radii are underestimated by factors $r/r_{\rm a} = 1.17$ if all planets orbit primaries, and $r/r_{\rm a} = 1.65$ if detected planets are equally likely to orbit primaries and singles. This approach makes minimal assumptions about the likelihood that a planet orbits the primary vs. the secondary (though Hirsch et al. 2017 do also consider weighting radius corrections by planet occurrence).

Our approach provides a suggestion of how to assess the probability that a detected planet in a binary system orbits either component. The problem requires assuming both the underlying radius distribution, and the relative number of planets per primary and secondary star. We found that for a broken power law radius distribution, with equal occurrence rates about primaries and secondaries, detected planets in binary systems are usually more likely to orbit the primary (Figure 5). For apparent radii greater than $2r_\oplus$, the odds of primary:secondary are 8:3. The odds increase for the planet to orbit the secondary increase at smaller apparent radii, and reach 1:1 at $r_{\rm a} = 0.4r_\oplus$. For *Kepler*, this suggests that for almost all confirmed KOIs in multiple star systems, the planet most likely orbits the primary.

Fraction of detected planets with binary companions—A related effect is that some detected planets are more likely to be in binary systems than others. Notably, Figure 5 predicts that the fraction of detected planets with binary companions should increase by 6 to 12% going from $2r_\oplus$ to $1.4r_\oplus$.

In Ziegler et al. (2017)'s recent summary of the Robo-AO KOI survey, they reported the fraction of planet-hosting stars with Robo-AO detected companion stars, binned over apparent planet radii. The fraction of KOIs with nearby stars and their 1σ uncertainties are:

- Earths ($r_{\rm a} < 1.6r_\oplus$): $16.3 \pm 1.0\%$, from 1480 systems.
- Neptunes ($1.6r_\oplus < r_{\rm a} < 3.9r_\oplus$): $13.0 \pm 0.8\%$, from 2058 systems.
- Saturns ($3.9r_\oplus < r_{\rm a} < 9r_\oplus$): $13.6 \pm 2.0\%$, from 338 systems.
- Jupiters ($r_{\rm a} > 9r_\oplus$): $19.0 \pm 2.8\%$, from 247 systems.

The uncertainties were calculated following Burgasser et al. (2003). The absolute values of the companion fractions are lower than in the solar neighborhood at least in part because of Robo-AO's sensitivity (Ziegler et al. 2017's Figure 2; Raghavan et al. 2010). The uptick in the companion rate for Jupiters is likely tied to a large astrophysical false positive rate (Santerne et al. 2012). Going from Neptunes to Earths, the data suggest a weak increase in the detected planet companion fraction, perhaps of a few percent. It will be interesting to see whether this effect is borne out by further observations.

Connection between high-resolution imaging and occurrence rates—Occurrence rate calculations are beginning to incorporate the results of high-resolution imaging surveys. For example, recent studies have reduced contamination in the numerator of the occurrence rate by using Furlan et al. (2017)'s catalog to test the effects of removing KOI hosts with known companions (Fulton et al. 2017; Petigura et al. 2017b). This is a good first step, but the denominator remains uncorrected.

Even if the multiplicity of every star that *Kepler* surveyed were known, the problem would not be solved. In this scenario, it would be possible to tabulate planet occurrence for single and binary star systems separately. The number of planets per single star would be known. However the number of planets per primary and secondary would still be convolved. An observational approach to separating the two populations might be to analyze the centroids, or the transit durations (and thus host star densities) for a representative sample of planets in binary systems.

A theory-driven solution might entail probabilistic population inference, with a parametrized model for planet populations around primary and secondary stars. Such an approach might be expected to find that the primary and secondary populations are different, since single M dwarfs host more small planets than FGK dwarfs (Dressing & Charbonneau 2015). Given the difficulty of writing a likelihood function in a survey with systematic biases, one possible approach might be to forward-model the survey in the framework of approximate Bayesian computation (*e.g.*, Morehead 2016).

6. CONCLUSION

This study presented a framework for estimating the impact of unresolved binaries on transit survey occurrence rates. From the order-of-magnitude argument presented in Section 3, we showed that for Sun-like stars in the local neighborhood, typical twin binary fractions yield apparent occurrence rates that are well within a factor of two of the true rates for single stars.

We then derived a general formula for the apparent rate density inferred by observers who ignore binarity (Equation 22). As input, this equation requires a volume-limited mass ratio distribution, and the true rate densities for planets around singles, primaries, and secondaries. The assumptions that enable this approach are as follows.

1. The transit survey is SNR-limited (Equation 1).

2. The observers know the true properties of the single and primary stars, and assign all unresolved binaries the properties of the primary.

3. There are two functions, $L(M)$ and $R(M)$, that specify a star's luminosity and radius in terms of its mass.

The interpretation of Equation 22 is that the apparent rate density is a weighted sum of the rate densities for single, primary, and secondary stars. The weights depend on the relative numbers of binary and single stars in the searchable volumes, which are affected by both dilution of transit signals and a Malmquist bias. Both effects should be considered in Monte Carlo simulations of transit surveys (*e.g.*, Bakos et al. 2013; Sullivan et al. 2015; Günther et al. 2017).

Applying Equation 22, we then showed that:

- If single, primary, and secondary stars all host the same number of planets per star, and the planet radius distribution is a power law with the exponent reported by Howard et al. (2012), then binarity influences apparent planet occurrence rates at the few percent level. This applies for apparent radii from $2r_\oplus$ to $17r_\oplus$ (Section 4.1).

- Assuming a broken-power law planet radius distribution, with the Howard et al. (2012) exponent above $2r_\oplus$ and a constant occurrence below $2r_\oplus$, the rate density for radii below $2r_\oplus$ is overestimated. The magnitude of the error, $\Delta\Gamma_0 = |\Gamma_0 - \Gamma_\mathrm{a}|/\Gamma_0$, can easily reach 50% (Figure 4). Although this is smaller than current systematic uncertainties on the occurrence rates of Earth-sized planets, this implies that binarity could eventually become an important component of the $\eta_\oplus$ error budget.

- Binarity blurs sharp features in the radius distribution (Figure 6), to a degree that could affect precise measurements of the depth, width, and slope of the Fulton et al. (2017) radius gap.

- Binarity does not lead to smaller apparent HJ occurrence rates (Figure 7). This assumes a Gaussian planet radius distribution, similar to that reported by Petigura et al. (2017b).

- Detected planets with apparent radii greater than $0.5r_\oplus$ that are revealed by high-resolution imaging surveys to exist in binaries are more likely to orbit the primary (Figure 5).
- Near the break in occurrence as a function of planet radius at $2r_\oplus$, the fraction of detected planets with binary companions should increase by roughly 5 to 10% (Figure 5).

These results are only strictly applicable for idealized transit surveys meeting the criteria mentioned above. For real transit surveys, although our approach is only suggestive, we hope that it provides a useful estimate for the systematic errors that can be incurred by ignoring binarity in calculations of planetary occurrence rates.

It was a pleasure discussing this study with T. Barclay, W. Bhatti, J. Christiansen, F. Dai, and T. Morton. This work made use of NASA's Astrophysics Data System Bibliographic Services.

Software: `numpy` (Walt et al. 2011), `scipy` (Jones et al. 2001), `matplotlib` (Hunter 2007), `pandas` (McKinney 2010), `IPython` (Pérez & Granger 2007)

APPENDIX

A. ALTERNATIVE DERIVATION OF APPARENT RATE DENSITY

This appendix derives Equation 22 through a more straight-forward, but also more tedious, procedure than that of Section 3.3. First, note that analogous to Equation 2, the apparent rate density $\Gamma_{\rm a}$ can be written

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{n_{\rm det}(r_{\rm a}, M_{\rm a})}{N_\star(r_{\rm a}, M_{\rm a})} \frac{1}{p_{\rm tra}(M_{\rm a})}, \tag{A1}$$

where $n_{\rm det}$ is the number of detected planets, per unit $(r_{\rm a}, M_{\rm a})$, with apparent radius $r_{\rm a}$ orbiting stars of apparent mass $M_{\rm a}$. The planets with $(r_{\rm a}, M_{\rm a})$ are associated with systems of many different planetary and stellar properties, so $n_{\rm det}(r_{\rm a}, M_{\rm a})$ is given by the convolution of the true rate density, $\Gamma(r, M)$, and $\mathcal{N}_\star(r_{\rm a}, M_{\rm a}; r, M)$, the number of searchable stars per unit $(r_{\rm a}, M_{\rm a})$ that give $(r_{\rm a}, M_{\rm a})$ when the true system actually has properties (r, M). Mathematically,

$$n_{\rm det}(r_{\rm a}, M_{\rm a}) = \sum_i n^i_{\rm det}(r_{\rm a}, M_{\rm a}) \tag{A2}$$

$$= \sum_i \int {\rm d}r{\rm d}M\, \mathcal{N}^i_\star(r_{\rm a}, M_{\rm a}; r, M)\, \Gamma_i(r, M)\, p_{\rm tra}(r, M), \tag{A3}$$

where i specifies the type of true host star (0: single, 1: primary, 2: secondary). The problem reduces to the evaluation of

$$\mathcal{N}^i_\star(\mathcal{P}_{\rm a}, \mathcal{S}_{\rm a}; \mathcal{P}, \mathcal{S}) \tag{A4}$$

for planets around single stars, primaries in binaries, and secondaries in binaries.

Single stars—For $i = 0$,

$$\mathcal{N}^0_\star(r_{\rm a}, M_{\rm a}; r, M) = \hat{\delta}(r_{\rm a} - r)\hat{\delta}(M_{\rm a} - M) N^0_\star(r, M), \tag{A5}$$

where $\hat{\delta}$ is the Dirac delta function, so

$$n^0_{\rm det}(r_{\rm a}, M_{\rm a}) = N^0_\star(r_{\rm a}, M_{\rm a})\, \Gamma_0(r_{\rm a}, M_{\rm a})\, p_{\rm tra}(r_{\rm a}, M_{\rm a}). \tag{A6}$$

Primaries in binaries—The number of primaries with apparent parameters $(r_{\rm a}, M_{\rm a})$ given the true parameters (r, M) is

$$\mathcal{N}^1_\star(r_{\rm a}, M_{\rm a}; r, M) = \int {\rm d}q\, f(q) \mathcal{N}^1_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M), \tag{A7}$$

where $f(q)$ is the volume-limited binary mass ratio distribution. Since we assume that binaries are assigned the mass of the primary,

$$\mathcal{N}^1_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M) \propto \hat{\delta}(M_{\rm a} - M). \tag{A8}$$

In this case, $\mathcal{N}^1_{\mathrm{s},q}$ is non-zero only at $r_\mathrm{a} = R_\mathrm{a}\sqrt{\delta}$, and the observed depth is

$$\delta = \left[\frac{r}{R(M_\mathrm{a})}\right]^2 \frac{L(M_\mathrm{a})}{L_\mathrm{tot}(M_\mathrm{a}, q)} \equiv \left[\frac{r}{R(M_\mathrm{a})}\right]^2 \frac{1}{\mathcal{D}_1^2} \tag{A9}$$

where

$$\mathcal{D}_1 = \sqrt{\frac{L_\mathrm{tot}(M_\mathrm{a}, q)}{L(M_\mathrm{a})}}. \tag{A10}$$

The normalization of $\mathcal{N}^1_{\mathrm{s},q}$ is given by the number of binaries that are searchable for a signal δ and that have mass ratio q:

$$N^0_\star(r_\mathrm{a}, M_\mathrm{a}) \, \frac{n_\mathrm{b}}{n_\mathrm{s}} \left[\frac{L_\mathrm{tot}(M_\mathrm{a}, q)}{L(M_\mathrm{a})}\right]^{3/2}. \tag{A11}$$

where the argument for the number of searchable single stars, $N^0_\star$, could also be expressed as $(\delta, L(M_\mathrm{a}))$. Applying Equation A11,

$$\mathcal{N}^1_{\mathrm{s},q}(r_\mathrm{a}, M_\mathrm{a}, q; r, M) = N^0_\star(r_\mathrm{a}, M_\mathrm{a}) \, \frac{n_\mathrm{b}}{n_\mathrm{s}} \, \mathcal{D}_1^3 \, \hat{\delta}\left(r_\mathrm{a} - \frac{r}{\mathcal{D}_1}\right) \hat{\delta}(M_\mathrm{a} - M). \tag{A12}$$

Secondaries in binaries—In this case, $M = qM_1 = qM_\mathrm{a}$, so

$$\mathcal{N}^2_{\mathrm{s},q}(r_\mathrm{a}, M_\mathrm{a}, q; r, M) \propto \hat{\delta}\left(M_\mathrm{a} - \frac{M}{q}\right). \tag{A13}$$

Again $\mathcal{N}^2_\star$ is non-zero only at $r_\mathrm{a} = R_\mathrm{a}\sqrt{\delta}$, but this time

$$\delta = \left[\frac{r}{R(qM_\mathrm{a})}\right]^2 \frac{L(qM_\mathrm{a})}{L_\mathrm{tot}(M_\mathrm{a}, q)} \equiv \left[\frac{r}{R(M_\mathrm{a})}\right]^2 \frac{1}{\mathcal{D}_2^2}, \tag{A14}$$

where

$$\mathcal{D}_2 = \frac{R(qM_\mathrm{a})}{R(M_\mathrm{a})}\sqrt{\frac{L_\mathrm{tot}(M_\mathrm{a}, q)}{L(qM_\mathrm{a})}}. \tag{A15}$$

The normalization remains the same as the previous case – we are counting the searchable stars at a given observed depth δ, and the total luminosity of the binary is the same. Thus,

$$\mathcal{N}^2_\star(r_\mathrm{a}, M_\mathrm{a}; r, M) = \int \mathrm{d}q \, f(q) \mathcal{N}^2_{\mathrm{s},q}(r_\mathrm{a}, M_\mathrm{a}, q; r, M), \tag{A16}$$

where

$$\mathcal{N}^2_{\mathrm{s},q}(r_\mathrm{a}, M_\mathrm{a}, q; r, M) = N^0_\star(r_\mathrm{a}, M_\mathrm{a}) \, \frac{n_\mathrm{b}}{n_\mathrm{s}} \, \mathcal{D}_1^3 \, \hat{\delta}\left(r_\mathrm{a} - \frac{r}{\mathcal{D}_2}\right) \hat{\delta}\left(M_\mathrm{a} - \frac{M}{q}\right). \tag{A17}$$

One might worry in Equation A17 that we opt to write $\mathcal{N}^2_\mathrm{s} \propto \hat{\delta}(M_\mathrm{a} - M/q)$, rather than $\propto \hat{\delta}(M_\mathrm{a}q - M)$ or some other delta function with the same functional dependence, but a different normalization once integrated. We do this because the delta function in Equation A17 is defined with respect to the measure $\mathrm{d}M_\mathrm{a}$, not $\mathrm{d}M$. This is because $\mathcal{N}^2_\mathrm{s}$ is defined as a number per r_a, per M_a.

Number of detected planets—Marginalizing per Equation A3, we find

$$n^0_{\rm det}(r_{\rm a}, M_{\rm a}) = \int {\rm d}r{\rm d}M\, \mathcal{N}^0_\star(r_{\rm a}, M_{\rm a}; r, M)\, \Gamma_0(r, M)\, p_{\rm tra}(M)$$
$$= N^0_\star(r_{\rm a}, M_{\rm a})\, \Gamma_0(r_{\rm a}, M_{\rm a})\, p_{\rm tra}(M_{\rm a}), \tag{A18}$$

and

$$n^1_{\rm det}(r_{\rm a}, M_{\rm a}) = \int {\rm d}r{\rm d}M\, \mathcal{N}^1_\star(r_{\rm a}, M_{\rm a}; r, M)\, \Gamma_1(r, M)\, p_{\rm tra}(M)$$
$$= N^0_\star(r_{\rm a}, M_{\rm a})\, p_{\rm tra}(M_{\rm a})\, \frac{n_{\rm b}}{n_{\rm s}} \int {\rm d}q\, \mathcal{D}_1^3 f(q)\, \mathcal{D}_1 \Gamma_1(\mathcal{D}_1 r_{\rm a}, M_{\rm a}). \tag{A19}$$

Finally,

$$n^2_{\rm det}(r_{\rm a}, M_{\rm a}) = \int {\rm d}r{\rm d}M\, \mathcal{N}^2_\star(r_{\rm a}, M_{\rm a}; r, M)\, \Gamma_2(r, M)\, p_{\rm tra}(M)$$
$$= N^0_\star(r_{\rm a}, M_{\rm a})\, \frac{n_{\rm b}}{n_{\rm s}} \int {\rm d}q\, \mathcal{D}_1^3 f(q)\, q\mathcal{D}_2 \Gamma_2(\mathcal{D}_2 r_{\rm a}, qM_{\rm a})\, \frac{R(qM_{\rm a})}{R(M_{\rm a})} q^{-1/3}. \tag{A20}$$

General formula for apparent occurrence rate—Combining the above results with Equation **??**, the apparent rate density,

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{N_\star(r_{\rm a}, M_{\rm a}) p_{\rm tra}(r_{\rm a}, M_{\rm a})} \sum_i n^i_{\rm det}(r_{\rm a}, M_{\rm a}), \tag{A21}$$

evaluates to

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{1+\mu} \left\{ \Gamma_0(r_{\rm a}, M_{\rm a}) + \frac{n_{\rm b}}{n_{\rm s}} \left[\int_0^1 {\rm d}q\, \mathcal{D}_1^3 f(q)\, \mathcal{D}_1 \Gamma_1(\mathcal{D}_1 r_{\rm a}, M_{\rm a}) \right.\right.$$
$$\left.\left. + \int_0^1 {\rm d}q\, \mathcal{D}_1^3 f(q)\, q\mathcal{D}_2 \Gamma_2(\mathcal{D}_2 r_{\rm a}, qM_{\rm a})\, \frac{R(qM_{\rm a})}{R(M_{\rm a})} q^{-1/3} \right] \right\}. \tag{A22}$$

We validate this equation in limits where it is possible to write down the answer (e.g., Equation 15), and also against a Monte Carlo realization of the twin binary models (Sections 3.1 and 4.1).

Table 1. Occurrence rates of hot Jupiters (HJs) about FGK dwarfs, as measured by radial velocity and transit surveys.

Reference	HJs per thousand stars	HJ Definition
Marcy et al. (2005)	12±2	$a < 0.1$ AU; $P \lesssim 10$ day
Cumming et al. (2008)	15±6	–
Mayor et al. (2011)	8.9±3.6	–
Wright et al. (2012)	12.0±3.8	–
Gould et al. (2006)	$3.1^{+4.3}_{-1.8}$	$P < 5$ day
Bayliss & Sackett (2011)	10^{+27}_{-8}	$P < 10$ day
Howard et al. (2012)	4±1	$P < 10$ day; $r_p = 8 - 32r_\oplus$; solar subset[a]
–	5±1	solar subset extended to $Kp < 16$
–	7.6±1.3	solar subset extended to $r_p > 5.6r_\oplus$.
Moutou et al. (2013)	10±3	*CoRoT* average; $P \lesssim 10$ day, $r_p > 4r_\oplus$
Petigura et al. (2017b)	$5.7^{+1.4}_{-1.2}$	$r_p = 8 - 24r_\oplus$; $P = 1 - 10$ day; CKS stars[b]
Santerne et al. (2018, in prep)	9.5±2.6	*CoRoT* galactic center
–	11.2±3.1	*CoRoT* anti-center

Note— The first four studies use data from radial velocity surveys; the rest are based on transit surveys. Many of these surveys selected different stellar samples. "–" denotes "same as above".

[a] Howard et al. (2012)'s "solar subset" was defined as *Kepler*-observed stars with $4100\,\mathrm{K} < T_{\mathrm{eff}} < 6100\,\mathrm{K}$, $Kp < 15$, $4.0 < \log g < 4.9$. They required signal to noise > 10 for planet detection.

[b] Petigura et al. (2017b)'s planet sample includes all KOIs with $Kp < 14.2$, with a statistically insignificant number of fainter stars with HZ planets and multiple transiting planets. Their stellar sample begins with Mathur et al. (2017)'s catalog of 199991 *Kepler*-observed stars. Successive cuts are: $Kp < 14.2$ mag, $T_{\mathrm{eff}} = 4700 - 6500\,\mathrm{K}$, and $\log g = 3.9 - 5.0$ dex, leaving 33020 stars.

REFERENCES

Adams, E. R., Ciardi, D. R., Dupree, A. K., et al. 2012, The Astronomical Journal, 144, 42

Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2013, The Astronomical Journal, 146, 9

Bakos, G. ., Csubry, Z., Penev, K., et al. 2013, PASP, 125, 154

Baranec, C., Ziegler, C., Law, N. M., et al. 2016, The Astronomical Journal, 152, 18

Bayliss, D. D. R., & Sackett, P. D. 2011, The Astrophysical Journal, 743, 103

Burgasser, A. J., Kirkpatrick, J. D., Reid, I. N., et al. 2003, The Astrophysical Journal, 586, 512

Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, The Astrophysical Journal, 809, 8

Cartier, K. M. S., Gilliland, R. L., Wright, J. T., & Ciardi, D. R. 2015, The Astrophysical Journal, 804, 97

Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, Publications of the Astronomical Society of the Pacific, 120, 531

Dong, S., & Zhu, Z. 2013, The Astrophysical Journal, 778, 53

Dong, S., Zheng, Z., Zhu, Z., et al. 2014, The Astrophysical Journal Letters, 789, L3

Dressing, C. D., Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2014, The Astronomical Journal, 148, 78

Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45

Everett, M. E., Barclay, T., Ciardi, D. R., et al. 2015, The Astronomical Journal, 149, 55

Fischer, D. A., & Valenti, J. 2005, The Astrophysical Journal, 622, 1102

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, The Astrophysical Journal, 795, 64

Fressin, F., Torres, G., Charbonneau, D., et al. 2013, The Astrophysical Journal, 766, 81

Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, The Astronomical Journal, 154, 109

Furlan, E., Ciardi, D. R., Everett, M. E., et al. 2017, The Astronomical Journal, 153, 71

Gilliland, R. L., Cartier, K. M. S., Adams, E. R., et al. 2015, The Astronomical Journal, 149, 24

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

Günther, M. N., Queloz, D., Demory, B.-O., & Bouchy, F. 2017, Monthly Notices of the Royal Astronomical Society, 465, 3379

Guo, X., Johnson, J. A., Mann, A. W., et al. 2017, The Astrophysical Journal, 838, 25

Hirsch, L. A., Ciardi, D. R., Howard, A. W., et al. 2017, The Astronomical Journal, 153, 117

Holman, M. J., & Wiegert, P. A. 1999, The Astronomical Journal, 117, 621

Horch, E. P., Howell, S. B., Everett, M. E., & Ciardi, D. R. 2012, The Astronomical Journal, 144, 165

—. 2014, The Astrophysical Journal, 795, 60

Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, The Astrophysical Journal Supplement Series, 201, 15

Howell, S. B., Everett, M. E., Sherry, W., Horch, E., & Ciardi, D. R. 2011, The Astronomical Journal, 142, 19

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90

Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, arXiv:1703.10402 [astro-ph]

Jones, E., Oliphant, T., Peterson, P., et al. 2001, Open source scientific tools for Python

Kraus, A. L., Ireland, M. J., Huber, D., Mann, A. W., & Dupuy, T. J. 2016, The Astronomical Journal, 152, 8

Law, N. M., Morton, T., Baranec, C., et al. 2014, The Astrophysical Journal, 791, 35

Lillo-Box, J., Barrado, D., & Bouy, H. 2012, Astronomy and Astrophysics, 546, A10

—. 2014, Astronomy and Astrophysics, 566, A103

Marcy, G., Butler, R. P., Fischer, D., et al. 2005, Progress of Theoretical Physics Supplement, 158, 24

Mathur, S., Huber, D., Batalha, N. M., et al. 2017, The Astrophysical Journal Supplement Series, 229, 30

Mayor, M., Marmier, M., Lovis, C., et al. 2011, ArXiv e-prints, 1109, arXiv:1109.2497

McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt & J. Millman, 51

Morehead, R. C. 2016, Ph.D. Thesis

Moutou, C., Deleuil, M., Guillot, T., et al. 2013, Icarus, 226, 1625

Owen, J. E., & Wu, Y. 2017, arXiv:1705.10810 [astro-ph]

Pepper, J., Gould, A., & Depoy, D. L. 2003, Acta Astronomica, 53, 213

Pérez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21

Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, Proceedings of the National Academy of Science, 110, 19273

Petigura, E. A., Marcy, G. W., & Howard, A. W. 2013, ApJ, 770, 69

Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017a, arXiv:1703.10400 [astro-ph]

Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2017b, arXiv:1712.04042 [astro-ph]

Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, The Astrophysical Journal Supplement Series, 190, 1

Santerne, A., Daz, R. F., Moutou, C., et al. 2012, Astronomy and Astrophysics, 545, A76

Santos, N. C., Israelian, G., & Mayor, M. 2004, Astronomy and Astrophysics, 415, 1153

Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, The Astrophysical Journal, 809, 77

Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2017, arXiv:1710.05398 [astro-ph]

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science & Engineering, 13, 22

Wang, J., Fischer, D. A., Horch, E. P., & Huang, X. 2015a, The Astrophysical Journal, 799, 229

Wang, J., Fischer, D. A., Horch, E. P., & Xie, J.-W. 2015b, The Astrophysical Journal, 806, 248

Wang, J., Fischer, D. A., Xie, J.-W., & Ciardi, D. R. 2015c, The Astrophysical Journal, 813, 130

Wang, J., Xie, J.-W., Barclay, T., & Fischer, D. A. 2014, The Astrophysical Journal, 783, 4

Winn, J. N. 2010, Exoplanet Transits and Occultations, ed. S. Seager (University of Arizona Press), 55

Wright, J. T., Marcy, G. W., Howard, A. W., et al. 2012, The Astrophysical Journal, 753, 160

Youdin, A. N. 2011, ApJ, 742, 38

Ziegler, C., Law, N. M., Baranec, C., et al. 2017, arXiv:1712.04454 [astro-ph]