Draft version January 30, 2018
Typeset using LaTeX **modern** style in AASTeX61

THE EFFECTS OF BINARITY ON PLANET OCCURRENCE RATES MEASURED BY TRANSIT SURVEYS

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 are actually binaries. Unresolved and unrecognized binary stars give rise to systematic errors in planet occurrence rates, including misclassified planets and miscounts in the number of searched stars. The individual errors can have different signs, making it difficult to anticipate the net effect on planet 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 $2r_\oplus$, the net systematic error is of order 10%. For smaller planets the errors are potentially larger: the occurrence of Earth-sized planets could be overestimated by as much as a factor of two. One consequence is that unrecognized binaries cannot account for the apparent discrepancy between the hot Jupiter occurrence rates measured by transit and RV surveys. We also demonstrate that if 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 about other worlds. One method for measuring occurrence rates is to monitor the brightnesses of many stars over a wide field, seeking evidence for planetary transits. Measuring occurrence rates using this method was the highest priority of the NASA *Kepler* mission. Great strides have been made in the analysis of *Kepler* data, including progress toward the goal of determining 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 in occurrence rate studies is that in most cases, investigators have assumed that all of the sources of light that were monitored are single stars (*e.g.*, Howard et al. 2012; Fressin et al. 2013; Foreman-Mackey et al. 2014; Dressing & Charbonneau 2015; Burke et al. 2015). In reality many of them are unresolved multiple-star systems, especially 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 "flux dilution" makes transit signals harder to detect and lowers the number of detections. On the other hand, in a binary there 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 was a serious issue at all, or even whether the net effect of all the errors is positive or negative. The goal of this study was to clarify the various sources of error and provide a framework for dealing with these issues. In this spirit, most of our models are idealized and analytic, and we do not attempt a detailed correction of the results from *Kepler* or any other real transit survey. However we can estimate the order of magnitude of the effects, assuming the survey is signal-to-noise limited and given a model for the true planet occurrence rate.

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 4, that allows for more realistic stellar and planetary populations. We use this formula in Section 5 to explore more complicated but nevertheless still analytic models. In Section 6 we discuss the errors due to unrecognized binaries for some specific cases of current interest: the occurrence of Earth-like planets; the apparent discrepancy between hot-Jupiter occurrence rates in different surveys; and the shape of the "evaporation valley" in the radius distribution of planets that was brought to light by Fulton et al. (2017). We review all the results in Section 7.

2. UNDERSTANDING THE ERRORS

Imagine a group of astronomers that wants to measure the mean number of planets of a certain type per star of a certain type. They perform a wide-field photometric search for planets that transit, and discover all for which

$$\frac{\delta_{\rm obs}}{\sigma} = ({\rm constant}) \cdot \delta_{\rm obs} L_{\rm sys}^{1/2} d^{-1} > \left(\frac{\rm S}{\rm N}\right)_{\rm min}. \tag{1}$$

Here the signal S is the observed transit depth $\delta_{\rm obs}$, in units of relative flux. The noise N is the uncertainty in the determination of the flux ratio of the star inside and outside of transits. Throughout this paper we will assume the photometric uncertainty to scale as the inverse square root of the number of photons collected from the source, *i.e.*, the dependence on stellar properties is $\sigma \propto (L_{\rm sys}/d^2)^{-1/2}$, where $L_{\rm sys}$ is the luminosity of the stellar system and d is its distance from Earth.

The astronomers then assume that all stars are single. In particular they do not have accurate enough parallaxes to tell that some of the stars are apparently overluminous. They want to compute the number of planets with size r per star of mass M. If the total number of such stars for which the planets could be detected ($N_\star$) and the number of detected planets ($N_{\rm det}$) are both large, then the astronomers' estimate for the occurrence rate (Λ) is simply

$$\Lambda(r, M) = \frac{N_{\rm det}(r, M)}{N_\star(r, M)} \times \frac{1}{p_{\rm tra}(r, M)}, \tag{2}$$

where $p_{\rm tra}$ is the geometric transit probability, needed to account 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 apparent transit 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.

We will focus exclusively on the subset of problems that arise from the fact that many stars exist in gravitationally bound binary systems. Even with this narrow focus, there are numerous sources of error. All three of the quantities in Equation 2 are subject to observational bias:

1. The number of detections, $N_{\rm det}$, 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 will 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 apparent number of stars that were searched, $N_\star$, is biased
   - toward lower values, because it does not include all of the secondary stars that were searched for transiting planets.
   - toward higher values, because some of the stars that appeared to be suitably small and bright to detect the desired type of planet are in fact binaries for which the amplitude of the photometric signal would be reduced to an undetectable level.
3. The transit probability $p_{\rm tra}$, which scales with the stellar mean density as $\rho^{-1/3}$, is biased because the planet-hosting star could be misclassified.

Two other problems that may arise are not represented in Equation 2. One is astrophysical: the true occurrence rate of the desired 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. 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, *e.g.*, Wang et al. 2015a).

The other problem is an observational effect. Even after the astronomers learn about binaries and attempt to correct for their presence, they must realize that the ratio of binaries to single stars in the sample will differ from that in a volume-limited sample due to a type of Malmquist bias. Binaries have a higher total luminosity than either the primary or secondary star would have on its own. This means that the binaries are searchable for transit signals of a given amplitude at greater distances from the Earth. The binaries in a sample of apparently searchable stars are therefore drawn from a larger search volume than the single stars.

Planet occurrence rates are often presented as two-dimensional density functions in the space of planet radius and orbital period. Thankfully none of the errors we have enumerated lead to errors in the period distribution. 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. Hence in what follows, we focus on the radius distribution and assume the periods are measured without significant uncertainty.

Given all of the confusing and opposing sources of error, we will proceed in stages. We will start from a barebones model in which 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. We consider only the corrections to the occurrence as a function of planetary radius, and not the orbital period (the other key variable in transit studies), because when multiple transits are detected the period can be directly measured regardless of whether the host star is single or part of a binary.

3. SIMPLE MODELS

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

Since the effects of binarity are most pronounced when the two components are similar, we begin by considering a universe in which all stars are identical, with mass M, radius R, and luminosity L. Some fraction of them exist in twin binaries. All planets have size $r_{\rm p}$. Our goal is to compare the occurrence inferred by observers who ignore binarity with the true occurrence in single star systems.

To clarify our terminology, we define an "apparent planet radius", $r_{\rm a}$, as the radius that one would infer from a transit signal under the assumption that the host star is single. The observed transit depth is given by the ratio of the apparent planet radius and the apparent stellar radius, $R_{\rm a}$:

$$\delta_{\rm obs} = \left(\frac{r_{\rm a}}{R_{\rm a}}\right)^2 = \left[\frac{r}{R_{\rm host}}\right]^2 \times \frac{L_{\rm host}}{L_{\rm sys}}, \tag{3}$$

where r is the planet radius, $R_{\rm host}$ ($L_{\rm host}$) is the host star's radius (luminosity), and $L_{\rm sys}$ is the entire system's luminosity. For the example at hand, $R_{\rm host} = R$, since all stars have the same size. In single star systems we have $r = r_{\rm a} = r_{\rm p}$. In twin binaries, $r_{\rm a} = r_{\rm p}/\sqrt{2}$, since $L_{\rm sys} = 2L$.

We can now understand that the observers are unwittingly reporting an *apparent* occurrence rate $\Lambda_{\rm a}(r_{\rm a})$ as a function of apparent radius. The apparent occurrence rate is the same as the *true* occurrence rate for transit signals of amplitude $\delta_{\rm obs} = (r_{\rm a}/R_{\rm a})^2$, *i.e.*, the number of detections ($N_{\rm det}$) of signals of amplitude $\delta_{\rm obs}$, divided by the number of sources $N_\star$ from which such a signal could be detected.

Finally, it is helpful to introduce an apparent rate density $\Gamma_{\rm a}$, the average number of planets per star per unit apparent planet radius. Analogous to Equation 2,

$$\Gamma_{\rm a}(r_{\rm a}) = \frac{n_{\rm det}(r_{\rm a})}{N_\star(r_{\rm a})} \times \frac{1}{p_{\rm tra}(r_{\rm a})}, \tag{4}$$

where $n_{\rm det} = {\rm d}N_{\rm det}/{\rm d}r_{\rm a}$ is the number of detected planets, per unit $r_{\rm a}$, with apparent radius $r_{\rm a}$. In Equation 4, we neglect the dependence on stellar mass since all stars in this model have the same properties.

To compute the number of detected planets, we consider the volumes within which a given signal could be detected. At a fixed signal-to-noise floor, there is a maximum searchable distance (Pepper et al. 2003; Pepper & Gaudi 2005). From Equation 1, this maximum searchable distance scales as

$$d \propto \delta_{\rm obs} \cdot L_{\rm sys}^{1/2}. \tag{5}$$

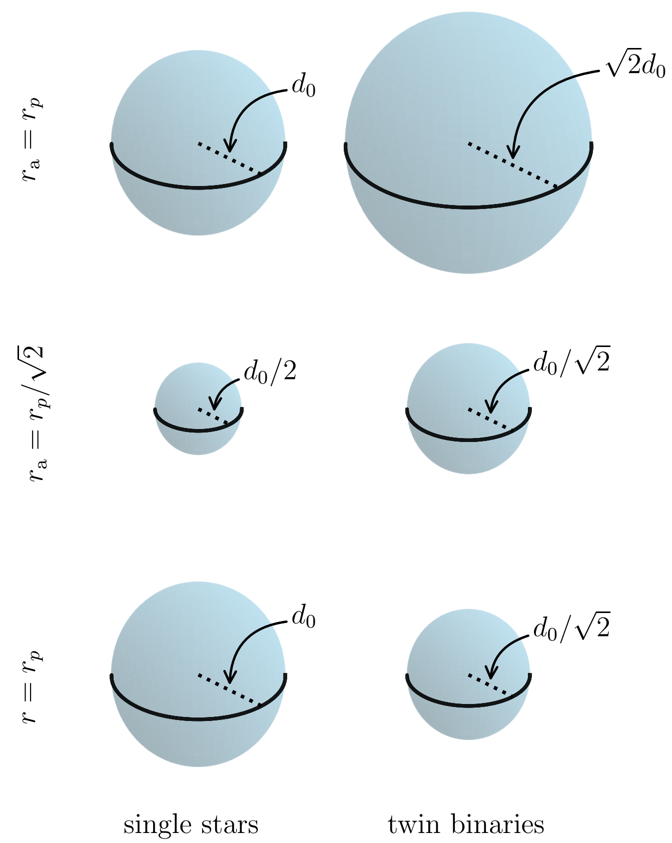

Figure 1. Searchable volumes for single stars (*left*) and twin binaries (*right*), assuming that all stars have the same mass, radius, and luminosity, and that all planets have the same radius $r = r_\mathrm{p}$. *Top:* At an apparent radius $r_\mathrm{a} = r_\mathrm{p}$, the observer searched twin binaries out to a distance $\sqrt{2}\times$ that of single stars. Because of dilution, there are no planets in twin binaries with $r_\mathrm{a} = r_\mathrm{p}$. *Middle:* At $r_\mathrm{a} = r_\mathrm{p}/\sqrt{2}$, the maximum searchable distances are half those at $r_\mathrm{a} = r_\mathrm{p}$. The only detected planets with $r_\mathrm{a} = r_\mathrm{p}/\sqrt{2}$ orbit twin binaries. *Bottom:* Planets with a true radius $r = r_\mathrm{p}$ are searchable to a maximum distance d_0 around singles, and $d_0/\sqrt{2}$ around twin binaries.

Assuming that stars are uniformly distributed in space, the number of searchable stars $N_\star$ is then proportional to

$$N_\star \propto n\delta_\mathrm{obs}^3 L_\mathrm{sys}^{3/2}, \tag{6}$$

where n is the number per unit volume of *e.g.*, single star, or binary systems. We neglect the dependence of n on stellar type.

The searchable volumes for this example are illustrated in Figure 1. At an apparent radius equal to the true radius ($r_\mathrm{a} = r_\mathrm{p}$), single stars with $d < d_0$ are searchable. Binaries with $r_\mathrm{a} = r_\mathrm{p}$ are also searchable, out to $\sqrt{2}d_0$. However, in this example no such systems exist; all planets in twin binaries have true radii $r = r_\mathrm{p}$, and so are observed with apparent radii $r_\mathrm{a} = r_\mathrm{p}/\sqrt{2}$, out to $d_0/\sqrt{2}$. At an apparent radius of

$r_{\rm a} = r_{\rm p}/\sqrt{2}$, there could also be planets with true radii $r = r_{\rm p}/\sqrt{2}$ orbiting singles, but there are no such systems in this model.

How many planets are detected? If we assume that there are $N_\star^0(r_{\rm a})$ singles that are searchable for planets with $r_{\rm a}$, and that there are Z_0 planets per single, then single stars contribute

$$n_{\rm det}^0(r_{\rm a}) = N_\star^0(r_{\rm a}) Z_0 p_{\rm tra} \cdot \delta(r_{\rm a} - r_p) \tag{7}$$

planet detections (per unit $r_{\rm a}$), for δ the Dirac delta function.

We can write similar equations for the binaries. First though, we define a ratio $\mu \equiv N_\star^{\rm b}(r_{\rm a})/N_\star^0(r_{\rm a})$ between the number of searchable binary systems and the number of searchable singles at any $r_{\rm a}$. This quantity is useful because although $N_\star^{\rm b}$ and $N_\star^0$ vary as a function of the observed transit depth, μ remains fixed. Applying this definition, at $r_{\rm a} = r_{\rm p}/\sqrt{2}$, some detections will come from primaries,

$$n_{\rm det}^1(r_{\rm a}) = \mu N_\star^0(r_{\rm a}) Z_1 p_{\rm tra} \cdot \delta\left(r_{\rm a} - \frac{r_{\rm p}}{\sqrt{2}}\right), \tag{8}$$

and some from secondaries,

$$n_{\rm det}^2(r_{\rm a}) = \mu N_\star^0(r_{\rm a}) Z_2 p_{\rm tra} \cdot \delta\left(r_{\rm a} - \frac{r_{\rm p}}{\sqrt{2}}\right), \tag{9}$$

where Z_1 is the number of planets per primary, Z_2 is the number of planets per secondary, and $p_{\rm tra}$ is the same for identical stars. For twin binaries, Equation 6 gives $\mu = 2^{3/2} n_{\rm b}/n_{\rm s}$, where $n_{\rm b}$ is the number density of binaries in a volume-limited sample.

The last item we need to write down the apparent rate density $\Gamma_{\rm a}(r_{\rm a})$ is the number of unresolved point-sources on sky that the observer thinks are searchable, $N_\star$. From the definition of μ,

$$N_\star(r_{\rm a}) = N_\star^0(r_{\rm a}) \times (1+\mu) \tag{10}$$

relates the number of apparently searchable point-sources to the number of searchable singles. From the discussion above, it should be clear that there are in fact $N_\star^0(r_{\rm a}) \times (1+2\mu)$ searchable stars at any given $r_{\rm a}$; the observers are under-counting.

Using Equation 4, the apparent rate density computed when ignoring binarity is

$$\Gamma_{\rm a}(r_{\rm a}) = \frac{1}{1+\mu} Z_0 \cdot \delta(r_{\rm a} - r_{\rm p}) + \frac{\mu}{1+\mu}(Z_1 + Z_2) \cdot \delta\left(r_{\rm a} - \frac{r_{\rm p}}{\sqrt{2}}\right). \tag{11}$$

The number of searchable singles at a given apparent radius, $N_\star^0(r_{\rm a})$, makes no appearance. The observer has erred in Equation 11 by under-counting the number of searchable stars at any apparent radius, and by inferring incorrect planet radii in binary systems.

To assess the severity of these errors, we must assume both a binary fraction, and also the true number of planets per single, primary, and secondary star (Z_0, Z_1, Z_2). For the former, Raghavan et al. (2010) found a multiplicity fraction of 0.44 for primaries with masses from $0.7 M_\odot$ to $1.3 M_\odot$. By multiplicity fraction, they meant

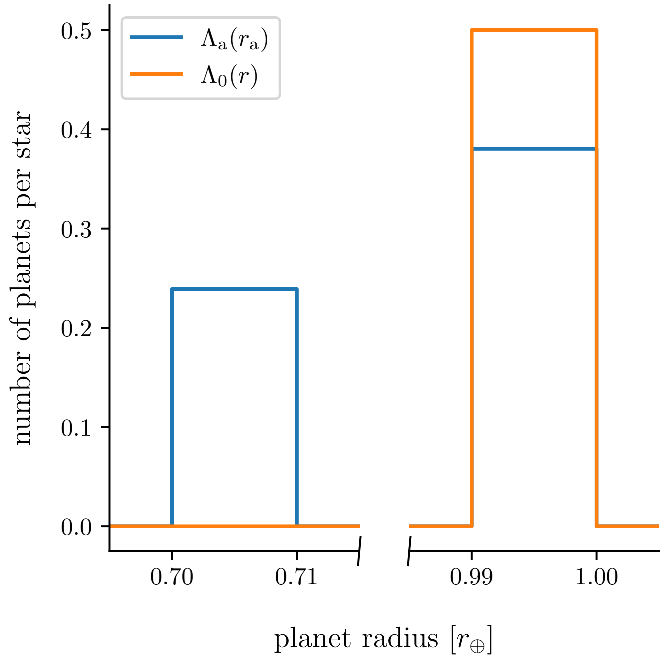

Figure 2. Apparent occurrence rate Λ_a, and occurrence rate for singles Λ_0, for a model with one type of star, one type of planet, and twin binaries (same as Figure 1). We assume a twin binary fraction of 0.1, and that single, primary, and secondary stars host planets at equal rates. If the true planet radius is r_p, all planets detected in binaries will have apparent radii $r_a = r_p/\sqrt{2}$; Equation 11 gives the normalizations.

"fraction of systems in a volume-limited sample that are multiple", which we simplistically take to be fraction of systems in a volume-limited sample that are binaries:

$$\text{binary fraction} \equiv \frac{n_b}{n_s + n_b} = 0.44, \tag{12}$$

Of course not all of these are binaries are close to being "twin" binaries as we have assumed in our simple calculation. Perhaps only a tenth of them have pairs of stars close enough in their basic properties to produce errors as significant as those we have been considering. Thus, an appropriate estimate for the binary fraction is of order 0.05. The ratio of the apparent to single star rate densities,

$$X_\Gamma(r) \equiv \left. \frac{\Gamma_a(r_a)}{\Gamma_0(r)} \right|_{r_a \to r}, \tag{13}$$

can then be thought of as a "correction factor". At the true planet radius, $X_\Gamma(r_p) = 1/(1+\mu)$. This yields a correction X_Γ of 0.87 (0.76) if the binary fraction is 0.05 (0.10). 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.

An alternative assumption is that secondaries do not host planets. In that case, $Z_0 = Z_1$, and $Z_2 = 0$. The correction to the rate density at the true planet radius becomes $X_\Gamma(r_p) = (1 + 2\mu)/(1 + \mu)^2$. This gives a correction of 0.98 (0.94) if the binary fraction is 0.05 (0.10). While the secondaries certainly do have planets, the fact that $\Gamma_{\rm a}/\Gamma_0$ is sensitive to the relative number of planets per single, primary, and secondary will become important as we proceed.

3.2. *One type of star, two types of planets*

A simple extension to the previous example will help us further distinguish the detected populations. Consider now a universe that is the same in every respect to that of Section 3.1, except that half of planets have true radii $r_{\rm p}$, while the other half have true radii $r_{\rm p}/\sqrt{2}$. The true rate densities are

$$\Gamma_i(r) = \frac{Z_i}{2}\left[\delta(r - r_{\rm p}) + \delta\left(r - \frac{r_{\rm p}}{\sqrt{2}}\right)\right], \quad \text{for } i \in \{0, 1, 2\}, \tag{14}$$

where as before $i = 0$ corresponds to singles, $i = 1$ to primaries, and $i = 2$ to secondaries, and the Z_i's are the number of planets per star of each type.

Following an identical line of reasoning as in Section 3.1, the apparent rate density can be written

$$\Gamma_{\rm a}(r_{\rm a}) = \frac{1}{2(1+\mu)}\left[Z_0 \cdot \delta(r_{\rm a} - r_{\rm p}) + [Z_0 + \mu(Z_1 + Z_2)] \cdot \delta\left(r_{\rm a} - \frac{r_{\rm p}}{\sqrt{2}}\right)\right.$$
$$\left. + \, \mu(Z_1 + Z_2) \cdot \delta\left(r_{\rm a} - \frac{r_{\rm p}}{2}\right)\right]. \tag{15}$$

The important qualitative point of this equation is that at $r_{\rm a} = r_{\rm p}/\sqrt{2}$, two populations contribute. The first population is singles with $r = r_{\rm p}/\sqrt{2}$, and the second is twin binaries with $r = r_{\rm p}$. Both populations are detected out to the same maximum searchable distance. Just as with Equation 10, we benefit from a cancellation of $N^0_\star(r_{\rm a})$ terms, leaving only the μ weights.

4. GENERAL FORMULA FOR APPARENT OCCURRENCE RATE

To generalize the procedure of Section 3, we consider an SNR-limited transit survey in which the singles and primaries can have arbitrary properties, but their true masses are known by the observer. We assume that there are some functions $L(M)$ and $R(M)$ that specify a star's luminosity and radius in terms of its mass, and that the volume-limited binary mass ratio distribution, $f(q)$, is independent of the primary's mass. We also allow arbitrary true rate densities for singles, primaries, and secondaries $(\Gamma_0, \Gamma_1, \Gamma_2)$. We presume that the observers always take the properties of a binary system to be those of the primary.

Given this laundry-list of assumptions, a general formula for the apparent rate density follows:

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{1 + \mu({\rm BF}, M_{\rm a})} \times \left\{\Gamma_0(r_{\rm a}, M_{\rm a}) + \frac{\rm BF}{1 - {\rm BF}}\left[\int_0^1 {\rm d}q\, \frac{f(q)}{\mathcal{A}^3} \cdot \frac{1}{\mathcal{A}}\Gamma_1\left(\frac{r_{\rm a}}{\mathcal{A}}, M_{\rm a}\right)\right.\right.$$
$$\left.\left. + \int_0^1 {\rm d}q\, \frac{f(q)}{\mathcal{A}^3} \cdot \frac{q}{\mathcal{B}}\Gamma_2\left(\frac{r_{\rm a}}{\mathcal{B}}, qM_{\rm a}\right) \cdot \frac{R(qM_{\rm a})}{R(M_{\rm a})}q^{-1/3}\right]\right\}, \tag{16}$$

where BF is the volume-limited binary fraction, and $\mathcal{A}(q, M_{\rm a})$ [$\mathcal{B}(q, M_{\rm a})$] are the dilution terms for primaries [secondaries], whose functional dependences are omitted for notational simplicity. The dilution terms can be written

$$\mathcal{A}(q, M_{\rm a}) = \sqrt{\frac{L(M_{\rm a})}{L_{\rm sys}(M_{\rm a}, q)}} = (1 + q^{\alpha})^{-1/2}, \tag{17}$$

and

$$\mathcal{B}(q, M_{\rm a}) = \frac{R(M_{\rm a})}{R(qM_{\rm a})}\sqrt{\frac{L(qM_{\rm a})}{L_{\rm sys}(M_{\rm a}, q)}} = q^{-1}(1 + q^{-\alpha})^{-1/2} \tag{18}$$

where the latter equalities assume $L \propto M^{\alpha} \propto R^{\alpha}$ (the former are more general). The rate densities depend on both apparent radius and mass because even if the primaries and singles were to have fixed mass, the mass of secondaries would vary, and so this dependence cannot be immediately omitted.

A derivation of Equation 16 is given in the Appendix; its qualitative features are as follows. The terms inside the curly braces $\{\ldots\}$ are the sum of the apparent rate densities from singles, primaries, and secondaries. The contribution from singles is not affected by binarity. In the contributions from primaries and secondaries, planets with $(r_{\rm a}, M_{\rm a})$ are associated with systems of different planetary and stellar properties, as determined by the mass ratio q of each binary. This corresponds to true planet radii and stellar masses $(r_{\rm a}/\mathcal{A}, M_{\rm a})$ for primaries, and $(r_{\rm a}/\mathcal{B}, qM_{\rm a})$ for secondaries. These contributions are then integrated over the mass ratio distribution including the effect of Malmquist bias, $f(q)/\mathcal{A}^3$, discussed further below. Finally, the latter term in the apparent rate density of secondaries comes from a correction for the overestimated transit probability.

Ratio of searchable binaries to singles—We must also specify the exact form of $\mu(\mathrm{BF}, M_{\rm a})$, the ratio of the number of searchable binary systems to singles. Given the observed signal $\delta_{\rm obs}$ and apparent stellar mass $M_{\rm a}$, recall that the maximum searchable distance for singles and binaries are proportional to $\delta_{\rm obs} \cdot L(M_{\rm a})^{1/2}$ and $\delta_{\rm obs} \cdot [L(M_{\rm a}) + L(qM_{\rm a})]^{1/2}$. Applying Equation 6,

$$\mu(\mathrm{BF}, M_{\rm a}) = \frac{N_{\star}^{\rm b}(r_{\rm a})}{N_s^0(r_{\rm a})} = \int_0^1 \frac{n_{\rm b}}{n_{\rm s}} \left[1 + \frac{L(qM_{\rm a})}{L(M_{\rm a})}\right]^{3/2} f(q)\,\mathrm{d}q. \tag{19}$$

If BF is independent of q, $L \propto M^{\alpha}$, and $f(q) \propto q^{\beta}$, this simplifies to

$$\mu = \frac{\mathrm{BF}}{1 - \mathrm{BF}} \cdot \frac{1}{1 + \beta} \int_0^1 (1 + q^{\alpha})^{3/2}\, q^{\beta}\,\mathrm{d}q, \tag{20}$$

which may be written in a closed form using the hypergeometric function.

This can be understood as a Malmquist bias: at fixed observed transit depth, for singles and primaries with the same luminosity, binaries are searchable to a greater

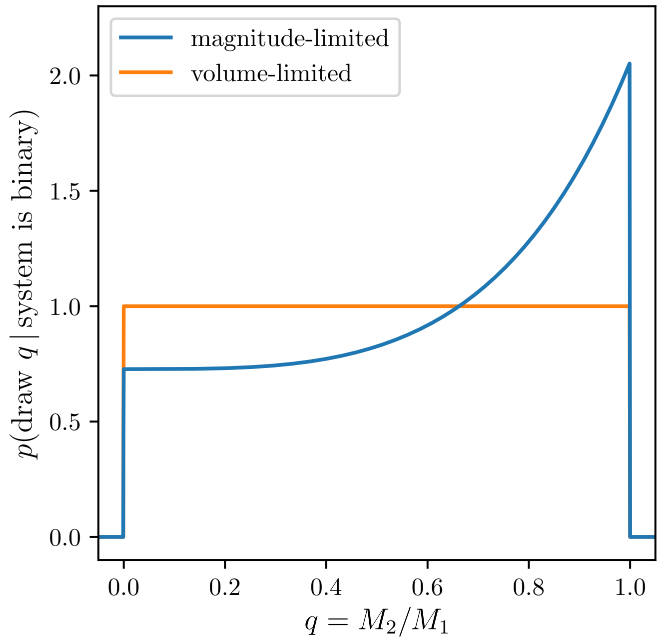

Figure 3. The mass ratio distribution for a magnitude-limited sample of binary stars, in which the underlying volume-limited distribution is uniform, qualitatively similar to Figure 16 of Raghavan et al. (2010). At a given observed transit depth, the searchable binaries in a transit survey are magnitude-limited.

distance; In particular, the sample of searchable binaries is magnitude-limited. Given a searchable binary, the probability that it has a mass ratio q scales as

$$p(\text{draw } q \,|\, \text{system is binary}) \propto (1+q^{\alpha})^{3/2} q^{\beta} \tag{21}$$

where q^{β} is the volume-limited probability of drawing a binary of mass ratio q, and the first term is the Malmquist bias when $L \propto M^{\alpha}$. We show this magnitude-limited mass ratio distribution for the $\beta = 0$ case in Figure 3. In Monte Carlo simulations of transit surveys, it is important to draw binaries from a correctly biased mass-ratio distribution (*e.g.*, Bakos et al. 2013; Sullivan et al. 2015; Günther et al. 2017).

5. REALISTIC STAR AND PLANET DISTRIBUTIONS

The following section applies our general equation for the apparent rate density (Equation 16) to study binarity's effects in regimes of observational interest. In general, we will write the rate density for each type of star, $\Gamma_i(r)$, as the product of a shape function and a constant:

$$\Gamma_i(r) = Z_i f_i(r), \quad \text{for } i \in \{0, 1, 2\}, \tag{22}$$

where as always, $i = 0$ corresponds to single stars, $i = 1$ to primaries of binaries, and $i = 2$ to secondaries of binaries. The shape function is normalized to unity. The Z_i's are each system type's occurrence rate Λ_i, integrated over all planetary radii. In other words, they are number of planets per single, primary, or secondary star. This section will consider the effects of varying both $f_i(r)$ and also the relative values of the Z_i's.

5.1. *Power law planet radius distributions*

5.1.1. *Twin binaries*

We begin introducing realism by keeping all binaries as twins, but letting the radius distribution of planets vary. Specifically, we take

$$\Gamma_i(r) = Z_i f(r) = Z_i r^\delta / \mathcal{N}_r, \tag{23}$$

for $f(r)$ the radius shape function, and $\mathcal{N}_r$ the shape function's normalization. The resulting apparent rate density is quite similar to that of the twin binary, fixed-planet case (Equation 11), except for a slightly different radius dependence and normalization:

$$\Gamma_\mathrm{a}(r_\mathrm{a}) = \frac{r_\mathrm{a}^\delta}{\mathcal{N}_r}\left[\frac{Z_0}{1+\mu} + 2^{\frac{\delta+1}{2}}\frac{\mu}{1+\mu}\left(Z_1 + Z_2\right)\right]. \tag{24}$$

As in Section 3.1, $\mu = 2^{3/2}\mathrm{BF}/(1-\mathrm{BF})$. If the Z_i's are equal, the "correction factor" relative to the rate density for singles becomes quite simple:

$$\left.\frac{\Gamma_\mathrm{a}(r_\mathrm{a})}{\Gamma_0(r)}\right|_{r_\mathrm{a}\to r} = \frac{1+2^{\frac{\delta+3}{2}}\mu}{1+\mu}. \tag{25}$$

For the case of $\mathrm{BF} = 0.1$, $\mu \approx 0.153$. Taking $\delta = -2.92$ from Howard et al. (2012), we get a correction factor $\Gamma_\mathrm{a}/\Gamma_0 = 1.004$. In other words, the apparent rate density is an *over*estimate compared to the rate density of single stars, with a relative difference $\delta\Gamma_0 = |\Gamma_0 - \Gamma_\mathrm{a}|/\Gamma_0$ of 0.4%. This is quite a small effect! Note though that it could change if the Z_i's are not equal. Also, the power law radius distribution $f(r) \propto r^\delta$ diverges for $\delta < 0$; Howard et al. (2012)'s results indicate that this approximation is good for $r_\mathrm{a} \gtrsim 2r_\oplus$. Similarly, it is nonsensical for $f(r)$ to be finite above some upper radius limit r_u, perhaps around $\approx 24r_\oplus$, based on the most inflated hot Jupiters. An upper cut-off of r_u would lead to smaller values of $\Gamma(r_\mathrm{a})$ down to $r_u/\sqrt{2}$, compared to an intrinsic power law without the cut-off. We are not particularly interested in this effect; there are better models for hot Jupiter occurrence rates that we will discuss in Section 5.3. If the assumptions behind this model (listed in Section 4) are at all applicable in real transit surveys, this suggests that for $2r_\oplus \lesssim r_\mathrm{a} \lesssim 17r_\oplus$, binarity's impact on derived occurrence rates is quite small.

5.1.2. *Binaries with power law mass ratio distribution*

To check whether non-twin binaries change the preceding result, we now assume $f(q) = q^\beta/\mathcal{N}_q$, for $\mathcal{N}_q$ the normalization. This changes μ (Equation 19 simplifies to Equation 20). It may also affect the rate density, which could be a function of the (varying) stellar mass. Absorbing this dependence into a power law as well,

$$\Gamma_i(r, M) = Z_i \times \frac{r^\delta}{\mathcal{N}_r} \times \frac{M^\gamma}{\mathcal{N}_M}, \tag{26}$$

where Z_i is dimensionless, and the normalization constants carry the units (each side has units $[r^{-1}M^{-1}]$). We assume that stars are a one-parameter family, given by $R \propto M \propto L^{\frac{1}{\alpha}}$, so that a given value of q determines everything about a secondary.

We are mostly interested in the apparent rate density's radius dependence. Marginalizing Equation 16 over apparent stellar mass, we find that when $Z_0 = Z_1 = Z_2$,

$$X_\Gamma \equiv \left. \frac{\Gamma_{\mathrm{a}}(r_{\mathrm{a}})}{\Gamma_0(r)} \right|_{r_{\mathrm{a}} \to r} \tag{27}$$

$$= \frac{1}{1+\mu} \left[1 + \frac{1}{\mathcal{N}_q} \frac{\mathrm{BF}}{1-\mathrm{BF}} \times \left(\int_0^1 \mathrm{d}q\, q^{\beta} (1+q^{\alpha})^{\frac{\delta+4}{2}} + \int_0^1 \mathrm{d}q\, q^{\beta+\delta+\frac{5}{3}} (1+q^{\alpha})^{\frac{3}{2}} (1+q^{-\alpha})^{\frac{\delta+1}{2}} \right) \right], \tag{28}$$

where γ does not appear because of the marginalization over M_{a}. For $\alpha = 3.5$, $\beta = 0$, $\gamma = 0$, $\delta = -2.92$, the summed integrals in Equation 28 give $(\ldots) \approx 1.50249$. For $\mathrm{BF} = 0.44$, this yields $\Gamma_{\mathrm{a}}/\Gamma_0 = 1.048$; the relative difference between the apparent rate density and the rate density around singles is 4.8%. This indicates that considering only twin binaries gave us correct intuition: for a power-law radius distribution ($2r_\oplus \lesssim r_{\mathrm{a}} \lesssim 17r_\oplus$) in which there are the same number of planets orbiting singles, primaries, and secondaries, binarity influences apparent planet occurrence rates around Sun-like stars at the ∼few percent level.

5.2. *Varying stars; broken power-law planet radius distribution*

Though the details of the true radius distribution $\Gamma_i(r)$ at $r < 2r_\oplus$ are currently an active topic of research, we can consider how binarity affects this regime under different plausible assumptions. For instance, assume that the true radius shape function is

$$f(r) \propto \begin{cases} r^{\delta} & \text{for } r \geq 2r_\oplus \\ \text{constant} & \text{for } r \leq 2r_\oplus. \end{cases} \tag{29}$$

As in the previous model, we assume that the observers know the true properties of all singles and primaries. The masses, radii, and luminosities of stars vary as $R \propto M \propto L^{\frac{1}{\alpha}}$. Our "nominal model" remains the same: $\alpha = 3.5$, $\beta = \gamma = 0$, $\delta = -2.92$.

At apparent radii $r_{\mathrm{a}} > 2r_\oplus$, the results of this model are the same as those from Section 5.1.2. For $r_{\mathrm{a}} < 2r_\oplus$, the equations are tedious, but still tractable. For simplicity, we insert Equation 29 into Equation 16, and integrate using a computer program. We refer the interested reader to our online implementation[1]. The output is validated against analytic predictions in the $r_{\mathrm{a}} > 2r_\oplus$ and the $r_{\mathrm{a}} < 2r_\oplus/\sqrt{2}$ regimes, and is plotted in Figure 4.

[1] https://github.com/lgbouma/binary_biases

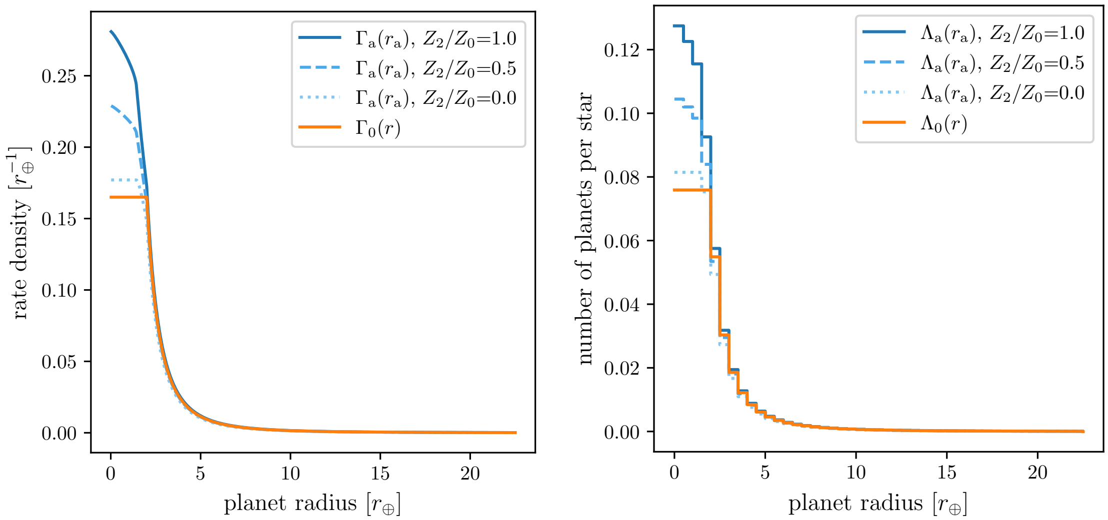

Figure 4. *Left:* apparent rate density (Γ_a) and single star rate density (Γ_0). *Right:* apparent occurrence rate (Λ_a) and single star occurrence rate (Λ_0), over $0.5r_\oplus$ bins. The true planet radius distribution is specified by Equation 29. This model assumes that the observer knows the true properties of all the singles and primaries, and that the volume-limited mass ratios of secondaries are drawn from a uniform distribution. Further, we take $Z_0 = Z_1 = 0.5$ throughout; Z_0, Z_1, Z_2 are the number of planets per single, primary, and secondary star. The rate and rate density are related by $\Lambda|_a^b = \int_a^b \Gamma\,\mathrm{d}r$.

The rate of Earth analogs—The immediately arresting result there is a "bump" in the apparent rate density at $r_a < 2r_\oplus$: the true rate for singles is less than the inferred rate. For the case in which secondaries host as many (half as many) planets as single stars, this means an overestimate of the absolute occurrence rate by $\approx 50\%$ ($\approx 25\%$). The "bump" exists even for the $Z_2/Z_0 = 0$ case as a $\approx 10\%$ effect. Evidently, the magnitude of this error depends strongly on the prevalence of planets around secondaries.

Hot Jupiter occurrence rates—Taking Figure 4 and integrating the rate density, we can compare the apparent hot Jupiter occurrence rate with the true rate for singles. In the most extreme case of $Z_2/Z_0 = 0$, we find that $\Lambda_{\mathrm{HJ},0}/\Lambda_{\mathrm{HJ,a}} = 1.13$, where

$$\Lambda_{\mathrm{HJ,a}} = \int_{8r_\oplus}^{\infty} \Gamma_a(r)\,\mathrm{d}r, \tag{30}$$

and similarly for $\Lambda_{\mathrm{HJ},0}$. If $Z_2/Z_0 > 0$, binarity affects the apparent hot Jupiter rate less: when $Z_2/Z_0 = 0.5$, we find $\Lambda_{\mathrm{HJ},0}/\Lambda_{\mathrm{HJ,a}} = 1.06$.

Fraction of detected planets from a given source—It would be nice to improve our intuition for how much the secondaries matter. We have the understanding that for a given true planet radius, secondaries have a much smaller searchable volume than primaries or singles. Does this necessarily imply that if we are given a detected planet with some apparent radius r_a, which is then observed by high-resolution imaging to exist in a binary, that the planet probably orbits the primary?

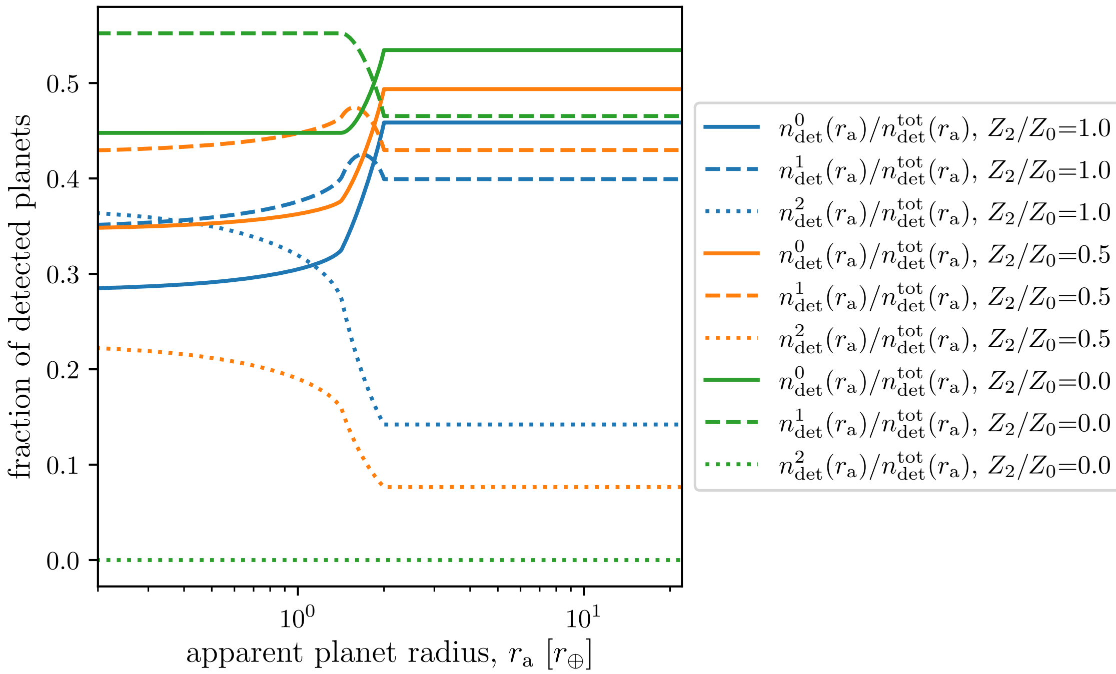

Figure 5. Fraction of detected planet at a given apparent planet radius coming from singles (solid lines), primaries (dashed lines), and secondaries (dotted lines). Three different values for Z_2/Z_0 are selected: 1 (blue), 0.5 (orange), 0 (green). The assumed true planet radius distribution is a broken power-law (Equation 29; same as Figure 4).

We explore this quantitatively in Figure 5. The lines in this figure are computed using the expressions given in the appendix for the number of detections coming from singles, primaries, and secondaries (Eqs. A17, A18, and A19).

In short, the detected planet usually is more likely to orbit the primary, but it depends on the number of planets orbiting secondaries, and also on the apparent radius of the detected planet. In the scenario that high-resolution imaging finds a system with $r_a > 2r_\oplus$ in a binary, roughly $\sim 2.5\times$ more detected planets orbit primaries than secondaries, for the $Z_2/Z_0 = 1$ case. For the $Z_2/Z_0 = 0.5$ case, $\sim 5\times$ more detected planets orbit primaries than secondaries. So detected planets in binaries with large apparent radii (relative to the cutoff in the intrinsic rate density) are more likely to orbit the primary.

The situation for $r_a < 2r_\oplus$ is more nuanced. For the $Z_2 = 0$ and $Z_2/Z_0 = 0.5$ cases, detected planets in binaries are still always more likely to orbit the primary. However, if there are as many planets orbiting the secondaries as the primaries, then there is a turn-over radius, $\approx 0.4r_\oplus$, beyond which more of the detected planets in binaries come from secondaries: they have all been "diluted down" from larger true planetary radii! Although this (apparent) radius is at the limit of current detection sensitivities, this effect will be interesting for future instruments to investigate.

Further, Figure 5 shows that going from apparent radii of $2r_\oplus$ to $\approx 1.4r_\oplus$, the fraction of detected planets with binary companions increases by $\approx 6-12\%$, de-

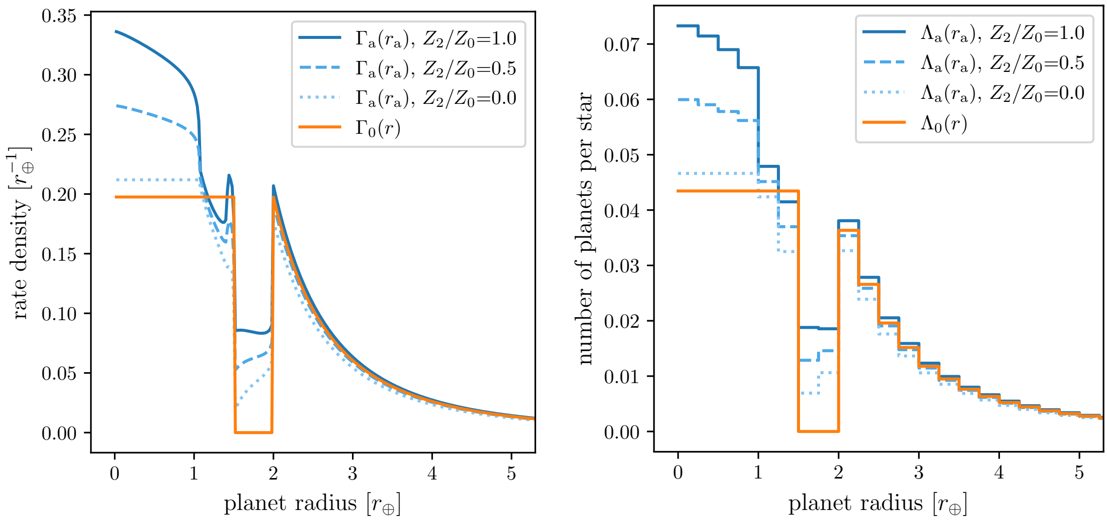

Figure 6. *Left:* rate density and *right:* rate (over $0.25r_\oplus$ bins) as a function of planet radius in a model with a radius gap (Equation 31). Other than the intrinsic radius distribution, this model has the same assumptions as Figure 4.

pending on the assumed value of Z_2/Z_0. Let the fraction of total detected planets from primaries at a given apparent radius be $F_1(r_a)$, so $F_1(r_a) \equiv n^1_{\rm det}(r_a)/n^{\rm tot}_{\rm det}$. Let the analogous quantity for secondaries be $F_2(r_a)$. In the planet-less secondaries case ($F_2(r_a) = 0$), one can analyze Eqs A17 and A18 semi-analytically and show that $F_1(2r_\oplus/\sqrt{2})/F_1(2r_\oplus) \approx 1.19$. Intuitively speaking, this "shift" is a feature of any radius distribution that declines at large radii, and flattens below some cutoff r_c. This is because (for dilution about primaries) at apparent radii $r_a < r_c/\sqrt{2}$, relatively more planets will be "diluted" into the given r_a, since there are more planets in the undiluted distribution from r_a to $\sqrt{2}r_a$. This prediction is compared against current measurements in Section 6.

5.3. *Further models: radius gaps, Gaussian HJ distributions*

The radius distribution specified by Equation 29 misses some important features recently reported by state of the art occurrence rate studies.

Precise features of the radius valley—In particular, Fulton et al. (2017) reported a "gap" in the radius-period plane (Petigura et al. 2017a; Johnson et al. 2017). The existence of the gap has been independently corroborated from a sample of KOIs with asteroseismically-determined stellar parameters (Van Eylen et al. 2017). Precise measurement of the gap's features, in particular its width, depth, and shape, will require accurate occurrence rates. To illustrate binarity's role in this problem, we make identical assumptions as in Section 5.2, but instead assume an intrinsic radius distribution

$$f(r) \propto \begin{cases} r^{\delta} & \text{for } r \geq 2r_{\oplus}, \\ 0 & \text{for } 1.5r_{\oplus} < r < 2r_{\oplus}, \\ \text{constant} & \text{for } r \leq 1.5r_{\oplus}. \end{cases} \tag{31}$$

The resulting true and inferred rates are shown in Figure 6. If left uncorrected, binarity makes the gap appear more shallow, and flattens the step-function edges. Of course, other effects could also "blur" the gap in the planet radius dimension. In particular, the valley's period-dependence is almost certainly not flat (Van Eylen et al. 2017; Owen & Wu 2017). This means that any study measuring the gap's width or depth in the face of binarity must either perform tests at fixed orbital period, or else marginalize over period and account for the associated blurring in the planet radius dimension.

Alternative models for the HJ distribution—In the recent work by Petigura et al. (2017b), hot Jupiters appear as an island in period-radius space, rather than as a continuous component of a power law distribution. This means that the apparent HJ rates computed in Section 5.2 are probably inaccurate. Instead, let us consider a Gaussian radius shape function,

$$f(r) \propto \exp\left(-\frac{(r-\bar{r})^2}{2\sigma^2}\right), \tag{32}$$

with $\bar{r} = 14r_{\oplus}$ and $\sigma = 2r_{\oplus}$. As always, $\Gamma_i(r) = Z_i f(r)$. We then compute $\Gamma_{\rm a}(r_{\rm a})$, and plot it in Figure 7. Integrating for $r_{\rm a} > 8r_{\oplus}$ to find hot Jupiter rates, we find the opposite effect as in the power law model. For $Z_2/Z_0 > 0$, the apparent HJ rate is *greater* than the true rate for singles. The effect is maximal when there are as many hot Jupiters orbiting secondaries as singles, in which case $\Lambda_{\rm HJ,0}/\Lambda_{\rm HJ,a} = 0.81$. For the case when no hot Jupiters orbit secondaries, $\Lambda_{\rm HJ,0} \approx \Lambda_{\rm HJ,a}$ to within one percent. Qualitatively, the apparent HJ rate can be greater than the true rate for singles because binary systems can have an extra star that yields hot Jupiter detections.

6. 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 stellar 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,

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

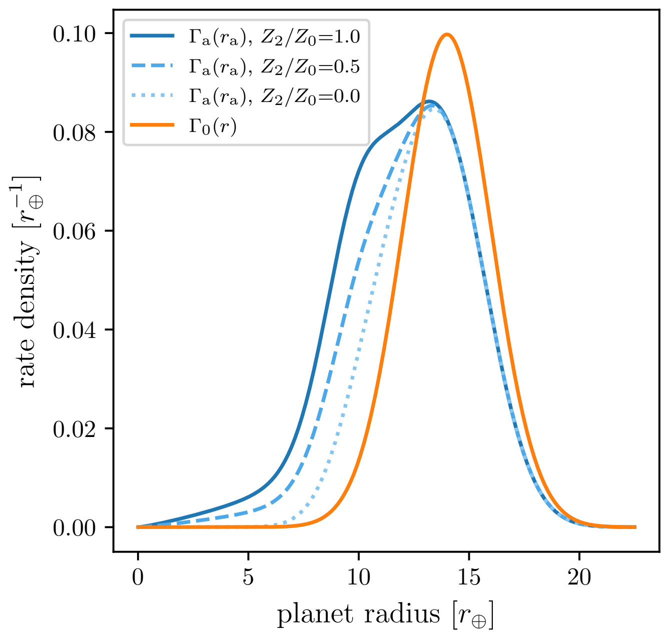

Figure 7. Rate density for a population of planets with true radii r drawn from a Gaussian with mean $14r_\oplus$ and standard deviation $2r_\oplus$. This is similar to the hot Jupiter distribution presented by Petigura et al. (2017b).

or lack thereof. Although this study does not resolve the problem, it does suggest the approximate scale of the possible errors in a survey-independent manner. The suggestion of Section 5.1 is that for $r_a \gtrsim 2r_\oplus$, binarity can be ignored down to the ~few percent precision level. For apparent radii $\lesssim 2r_\oplus$, the picture is less forgiving: Section 5.2 suggests that relative errors between apparent rates and true rates around singles could easily reach $\sim$50%.

The rate of Earth analogs—In line with *Kepler*'s primary science goal, the rate of Earth-like planets orbiting Sun-like stars has been independently measured by Youdin (2011); Petigura et al. (2013); Dong & Zhu (2013); Foreman-Mackey et al. (2014), and Burke et al. (2015). These efforts found that the one-year terrestrial planet occurrence rate is between $\approx$0.03 and $\approx$1 per Sun-like star, depending on assumptions that are made when calculating the rate (see Burke et al. 2015, , Figure 17). In Section 5.2, our power law model indicates that when primaries, singles, and secondaries host the same number of planets, the apparent rate is 50% higher than the true rate around single stars. Though this bias seems large, it is currently smaller than the other systematic factors that 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 to this assessment of binarity's importance for $\eta_\oplus$ measurements is that none of our models included the rate density's period-dependence. However, close binaries usually provoke dynamical instabilities, leading to fewer long-period planets per star (*e.g.*, Holman & Wiegert 1999; Wang et al. 2014; Kraus et al. 2016). This might affect transit survey measurements of $\eta_\oplus$ beyond our rough estimate.

Hot Jupiter rate discrepancy—While binarity may appreciably affect $\eta_\oplus$ measurements, our analysis suggests that it is unlikely to influence the hot Jupiter rate discrepancy. The context of this disagreement is that hot Jupiter occurrence rates measured by transit surveys ($\approx 0.5\%$) are marginally lower than those found by radial velocity surveys ($\approx 1\%$; see Table 1). Though the discrepancy has weak statistical significance ($< 3\sigma$), one reason to expect a difference is that the corresponding stellar populations have distinct metallicities. As argued by Gould et al. (2006), the RV sample is biased towards metal-rich stars, which have been measured by RV surveys to preferentially host more giant planets (Santos et al. 2004; Fischer & Valenti 2005). 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 measured 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 this would imply that the metallicity difference could account for a $\approx 20\%$ relative difference in the measured rates between the CKS and *Kepler* samples – not a factor of two[3]. Guo et al. concluded that "other factors, such as binary contamination and imperfect stellar properties" must also be at play.

The hypothesis that binarity might matter is grounded in the fact that radial velocity and transit surveys treat binarity differently. Radial velocity surveys typically reject both visual and spectroscopic binaries (*e.g.*, Wright et al. 2012). Transit surveys observe binaries, but the question of whether they were searchable to begin with is left for later interpretation. In spectroscopic follow-up of transiting planet candidates, the prevalence of astrophysical false-positives may also lead to a tendency against confirming transiting planets in binary systems. A separate observational bias is that the RV surveys typically observe chromospherically "quiet" stars, which may bias their detection probability high (Bastien et al. 2014).

Ignoring these complications, in this work we concentrated on binarity's effects on star counts and the apparent radii of detected planets. Assuming a power-law radius distribution, and that no secondaries host HJs, our results from Section 5.2 indicate that binarity could lead to underestimated HJ rates relative to singles by a multiplicative factor of at most ≈ 1.13. We later pointed out in Section 5.3 that assuming a power-law radius distribution is probably wrong if one wishes to study the HJ rate discrepancy. If we instead assumed a Gaussian radius distribution (following Petigura et al. 2017b), apparent HJ rates are *greater* than the true rate around singles: the effect goes the wrong way. The "hot Jupiter island" reported by Petigura et al.

[3] Petigura et al. (2017b) recently found $\beta = 3.4^{0.9}_{-0.8}$. This gives a $\approx 25\%$ relative difference, indicating metallicity could account for about half of the "hot Jupiter rate discrepancy".

(2017b) cannot possibly be part of a continuous power law. Thus binarity is unlikely to resolve the HJ rate discrepancy.

Of course, such calculations are only suggestive; a conclusive resolution of binarity's effects may require a detailed understanding of the *Kepler* field's multiplicity statistics, and the mission's completeness (both for candidate detection and follow-up). For instance, if hot Jupiters are less likely to be confirmed in binary systems, this might bias the rates low. Alternatively, *TESS* is expected to discover over 10^4 giant planets (Ricker et al. 2014; Sullivan et al. 2015). Though they will be difficult to distinguish from false positives, one possible use of this sample will be to measure an occurrence rate of short-period giant planets with minimal error from counting statistics. Our models suggest that binarity will only become an appreciable fraction of the error budget at the $\approx 10\%$ precision level. This should be sufficient for a rate measurement precise enough to indicate a preference between $\approx 0.5\%$ and $\approx 1\%$ of Sun-like stars hosting hot Jupiters.

Practical effects: smooth detection efficiencies; finite SNR floors—To apply this study to *Kepler* and other real transit surveys, a few practical concerns become pressing. One concern is that space-based missions with telemetric requirements are often required to select target stars. Even if the selection process is aimed at creating an SNR-limited survey, as Batalha et al. (2010) did for *Kepler*, uncertainties about stellar parameters and on-sky instrument performance can lead to deviations away from an SNR-limited stellar sample.

Further, in most transit surveys the detection efficiency is not a step function in SNR. Transit survey pipeline detection efficiencies are usually smooth functions in SNR; see for example the injection & recovery simulations performed by Christiansen et al. (2016) on *Kepler* data, and the resulting smooth functional form adopted by Fulton et al. (2017). If we were to take a smooth detection probability in SNR, we would need to include a detection efficiency term in Equation 4 to inversely weight for planets with low probabilities of being detected.

The simplest way to avoid the conceptual complications of accounting for finite detection efficiency is to raise the SNR floor, to *e.g.,* SNR ≈ 12 for *Kepler* (Fulton et al. 2017, Figure 5). To a good approximation, this allows one to use the boolean distinctions "searchable" and "not searchable" for signals of an observed depth (at fixed orbital period). This also makes the survey strictly SNR-limited, rather than "fuzzily SNR-limited". Since the minimum SNR for detection is set in an arbitrary manner anyway[4], this simplification would put occurrence rate studies on clearer conceptual ground, and facilitate the process of converting the observed distributions of apparent radii into true radius distributions.

[4] For example, *Kepler*'s threshold of SNR $= 7.1$ was set for there to be no more than one statistical false alarm across the full *Kepler* campaign (Jenkins et al. 2002). Equally valid would have been to insist on no more than 10^{-3} false alarms per survey, raising the detection floor.

Finally, our calculations have ignored the fact that any given transit survey's finite SNR floor might censor the apparent rate density. If the surveyed stars all have the same size, this is equivalent to saying that there might be a cutoff in apparent planet radii, below which no planets are detected. The effects on our main results (e.g., Figs. 4, 5, and 7) would be that the apparent rate density below the cutoff in apparent radius would simply drop to zero.

Connecting high resolution imaging to occurrence rates—A direct reason to address stellar multiplicity in transit surveys is that it changes the interpretation of planet candidates on a system-by-system level. Consequently, high resolution imaging campaigns have measured the multiplicities of 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 collected by Furlan et al. (2017), and they represent an important advance in understanding the KOI sample's multiplicity statistics. In particular, they can be immediately applied to rectify binarity's effects on the mass-radius diagram (Furlan & Howell 2017).

These high resolution imaging campaigns are also beginning to connect with occurrence rate calculations. The most recent rate studies have used Furlan et al. (2017)'s catalog to test the effects of removing KOI hosts with known companions, which helps reduce contamination in the "numerator" of the occurrence rate (Fulton et al. 2017; Petigura et al. 2017b). However, without an understanding of the multiplicity statistics of the non-KOI stars, the true number of searchable stars, and thus the true occurrence rates, will remain biased.

Does a detected planet orbit the primary or secondary?—One important step towards rectifying binarity's effects is identifying the host star for any detected planets in binaries. The current wisdom is that nothing substantive can be said about the likelihood of a planet orbiting the primary vs. the secondary (*e.g.*, Ciardi et al. 2015; Ziegler et al. 2017). This of course ignores cases when planets can be confirmed to orbit either component by analyzing the transit durations or centroids.

Assuming a broken power-law radius distribution, we showed in Figure 5 that there are many cases where a detected planet is much more likely to orbit the primary. A planet orbiting the secondary does lead to extreme corrections. However, at $r_a \gtrsim 2r_\oplus$, these cases are rare outliers. At $r_\oplus \lesssim r_a \lesssim 2r_\oplus$, for all of the trial values of Z_2/Z_0 assumed in Figure 5, a detected planet is always more likely to orbit the primary than the secondary. Only at very small apparent radii, $r_a \approx 0.5r_\oplus$, and only if secondaries host as many planets as singles and primaries, does this "rule" break down.

Of course, our model's assumptions (see the list in Section 7) might not apply to the *Kepler* dataset. However if they are applicable, then for the 56% of "`CONFIRMED`"

KOIs[5] with apparent radii $> 2r_\oplus$, whenever high-resolution imaging discovers a binary companion in a system that hosts a detected transiting planet, the planet is much more likely to orbit the primary.

Fraction of detected planets with binary companions vs. apparent radius—One further prediction of Figure 5 is that the fraction of detected planets with binary companions increases by $\approx 6-12\%$ going from $2r_\oplus$ to $\approx 1.4r_\oplus$. 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, binning by Earths ($r_\mathrm{a} < 1.6r_\oplus$), Neptunes ($1.6r_\oplus < r_\mathrm{a} < 3.9r_\oplus$), Saturns ($3.9r_\oplus < r_\mathrm{a} < 9r_\oplus$), and Jupiters ($r_\mathrm{a} > 9r_\oplus$). The reported nearby star rates and their 1σ uncertainties are:

- Earths: $16.3 \pm 1.0\%$, from 1480 systems.
- Neptunes: $13.0 \pm 0.8\%$, from 2058 systems.
- Saturns: $13.6 \pm 2.0\%$, from 338 systems.
- Jupiters: $19.0 \pm 2.8\%$, from 247 systems.

where the uncertainties were calculated following Burgasser et al. (2003). The absolute values of the rates are less than the $\sim 45\%$ typical for Sun-like stars at least in part because of Robo-AO's sensitivity (Ziegler et al. 2017's Figure 2; Raghavan et al. 2010). Another reason for the lower measured companion rates could be that planetary systems are less likely to have binary companions (as Kraus et al. 2016 argued for close binary companions). 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.

7. CONCLUSION

This study presented simple models for how binarity affects occurrence rates measured by transit surveys. The simplest of these models (Section 3) provided an order-of-magnitude estimate that suggested we would need implausibly high twin binary fractions for binarity to affect apparent occurrence rates by more than factors of two.

We then derived a general formula for the apparent rate density inferred by an observer ignoring binarity (Equation 16). As input, this equation requires a volume-limited mass ratio distribution $f(q)$, and the true rate densities for planets around singles, primaries, and secondaries. The assumptions that make the model tractable are:

1. the transit survey is SNR-limited;

[5] Exoplanet Archive; Akeson et al. (2013); `exoplanetarchive.ipac.caltech.edu`

2. the true properties of all the singles and primaries are known to the observer;
3. the observers assume that binaries have the same properties as the primaries;
4. there are functions $L(M)$ and $R(M)$ that specify a star's luminosity and radius in terms of its mass.

Allowing for these conditions, many results follow:

- Assuming a power-law planet radius distribution, and that the same number of planets orbit singles, primaries, and secondaries, binarity influences apparent planet occurrence rates around Sun-like stars at the ∼few percent level from radii $2r_\oplus \lesssim r_\mathrm{a} \lesssim 17r_\oplus$ (Section 5.1).
- Assuming a broken-power law planet radius distribution, with Howard et al. (2012)'s exponent at $r > 2r_\oplus$ and a constant occurrence at $r < 2r_\oplus$, there is a "bump" in the apparent rate density at $r_\mathrm{a} < 2r_\oplus$, leading to a relative error $\delta\Gamma_0 = |\Gamma_0 - \Gamma_\mathrm{a}|/\Gamma_0$ of at most $\approx 50\%$ (Figure 4). Although this is smaller than current systematic uncertainties on the occurrence rates of Earth-sized planets, this means that binarity could eventually become an important component of the $\eta_\oplus$ error budget.
- Binarity "fills in" gaps in the radius distribution (Figure 6), by an amount that could affect precise measurements of the depth, width, and slope of Fulton et al. (2017)'s radius gap, in the event that planets are not carefully vetted with high resolution imaging.
- 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 $r_\mathrm{a} \gtrsim 0.5r_\oplus$ 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 the rate distribution ($\approx 2r_\oplus$), the fraction of detected planets with binary companions increases by $\approx 5 - 10\%$ (Figure 5).

All of these results should be understood as only being strictly applicable when the assumptions listed above are met. Otherwise, while it is only suggestive, our approach still provides helpful hints at how ignoring stellar binarity influences transit survey occurrence rates.

It was a pleasure sharing discussions about this study with T. Barclay, W. Bhatti, and F. Dai. 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)

APPENDIX

A. DERIVATION OF GENERAL FORMULA FOR APPARENT OCCURRENCE RATE

In this appendix, we derive Equation 16. First, recall our definition of apparent rate density (Equation 4). We assume that R and L are uniquely determined from the assumed stellar mass–radius–luminosity relation, and neglect the dependence on planetary orbital period. 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 (per unit $(r_{\rm a}, M_{\rm a})$) of searchable stars that give $(r_{\rm a}, M_{\rm a})$ when the true system actually has (r, M). Mathematically,

$$n_{\rm det}(r_{\rm a}, M_{\rm a}) = \sum_i n^i_{\rm det}(r_{\rm a}, M_{\rm a}) \tag{A1}$$

$$= \sum_i \int {\rm d}r{\rm d}M\, \mathcal{N}^i_\star(r_{\rm a}, M_{\rm a}; r, M) \cdot \Gamma_i(r, M) \cdot p_{\rm tra}(r, M), \tag{A2}$$

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

$$\mathcal{N}^i_\star(\mathcal{P}_{\rm a}, \mathcal{S}_{\rm a}; \mathcal{P}, \mathcal{S}) \tag{A3}$$

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

Single stars—For $i = 0$,

$$\mathcal{N}^0_\star(r_{\rm a}, M_{\rm a}; r, M) = \delta(r_{\rm a} - r)\delta(M_{\rm a} - M)N^0_\star(r, M), \tag{A4}$$

so

$$n^0_{\rm det}(r_{\rm a}, M_{\rm a}) = N^0_\star(r_{\rm a}, M_{\rm a}) \cdot \Gamma_0(r_{\rm a}, M_{\rm a}) \cdot p_{\rm tra}(r_{\rm a}, M_{\rm a}). \tag{A5}$$

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

$$\mathcal{N}^1_\star(r_{\rm a}, M_{\rm a}; r, M) = \int {\rm d}q\, f(q)\mathcal{N}^1_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M), \tag{A6}$$

where $f(q)$ is the volume-limited binary mass ratio distribution.

Since we assume $\mathcal{S}_{\rm a} = \mathcal{S}_1$,

$$\mathcal{N}^1_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M) \propto \delta(M_{\rm a} - M). \tag{A7}$$

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

$$\delta_{\rm obs} = \left[\frac{r}{R(M_{\rm a})}\right]^2 \times \frac{L(M_{\rm a})}{L_{\rm sys}(M_{\rm a}, q)} \equiv \left[\frac{r}{R(M_{\rm a})}\right]^2 \times \mathcal{A}(q, M_{\rm a})^2, \tag{A8}$$

where

$$\mathcal{A}(q, M_{\rm a}) = \sqrt{\frac{L(M_{\rm a})}{L_{\rm sys}(M_{\rm a}, q)}}. \tag{A9}$$

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

$$N^0_\star(\delta_{\rm obs}, L(M_{\rm a})) \cdot \frac{n_b}{n_s} \left[\frac{L_{\rm sys}(M_{\rm a}, q)}{L(M_{\rm a})}\right]^{3/2} = N^0_\star(\delta_{\rm obs}, L(M_{\rm a})) \cdot \frac{\rm BF}{1-{\rm BF}} \cdot \frac{1}{\mathcal{A}(q, M_{\rm a})^3}. \tag{A10}$$

Thus,

$$\begin{aligned}\mathcal{N}^1_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M) = N^0_\star(\delta_{\rm obs}, L(M_{\rm a})) \cdot \frac{\rm BF}{1-{\rm BF}} \cdot \frac{1}{\mathcal{A}(q, M_{\rm a})^3} \\ \times \delta\left[r_{\rm a} - r\mathcal{A}(q, M_{\rm a})\right] \delta(M_{\rm a} - M).\end{aligned} \tag{A11}$$

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

$$\mathcal{N}^2_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M) \propto \delta\left(M_{\rm a} - \frac{M}{q}\right). \tag{A12}$$

Again $\mathcal{N}^2_\star$ is non-zero only at $r_{\rm a} = R_a\sqrt{\delta_{\rm obs}}$, but this time

$$\delta_{\rm obs} = \left[\frac{r}{R(qM_{\rm a})}\right]^2 \times \frac{L(qM_{\rm a})}{L_{\rm sys}(M_{\rm a}, q)} \equiv \left[\frac{r}{R(M_{\rm a})}\right]^2 \times \mathcal{B}(q, M_{\rm a})^2, \tag{A13}$$

where

$$\mathcal{B}(q, M_{\rm a}) = \frac{R(M_{\rm a})}{R(qM_{\rm a})}\sqrt{\frac{L(qM_{\rm a})}{L_{\rm sys}(M_{\rm a}, q)}}. \tag{A14}$$

The normalization remains the same as the previous case (we are counting the searchable stars at a given observed depth $\delta_{\rm obs}$, and the total luminosity of the binary is the same). Thus,

$$\mathcal{N}^2_\star(r_{\rm a}, M_{\rm a}; r, M) = \int {\rm d}q\, f(q) \mathcal{N}^2_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M), \tag{A15}$$

where

$$\begin{aligned}\mathcal{N}^2_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M) = N^0_\star(\delta_{\rm obs}, L(M_{\rm a})) \cdot \frac{\rm BF}{1-{\rm BF}} \cdot \frac{1}{\mathcal{A}(q, M_{\rm a})^3} \\ \times \delta\left[r_{\rm a} - r\mathcal{B}(q, M_{\rm a})\right] \delta\left(M_{\rm a} - \frac{M}{q}\right).\end{aligned} \tag{A16}$$

One might worry in Equation A16 that we opt to write $\mathcal{N}^2_{\rm s} \propto \delta(M_{\rm a} - M/q)$, rather than $\propto \delta(M_{\rm a}q - M)$ 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 A16 is defined with respect to the measure ${\rm d}M_{\rm a}$, not ${\rm d}M$. This is because $\mathcal{N}^2_{\rm s}$ is defined as a number per $r_{\rm a}$, per $M_{\rm a}$.

Number of detected planets—Marginalizing per Equation A2, we find

$$\begin{aligned} n^0_{\rm det}(r_{\rm a}, M_{\rm a}) &= \int {\rm d}r{\rm d}M\, \mathcal{N}^0_\star(r_{\rm a}, M_{\rm a}; r, M) \cdot \Gamma_0(r, M) \cdot p_{\rm tra}(M) \\ &= N^0_\star(\delta_{\rm obs}, L(M_{\rm a})) \cdot \Gamma_0(r_{\rm a}, M_{\rm a}) \cdot p_{\rm tra}(M_{\rm a}), \end{aligned} \tag{A17}$$

and

$$\begin{aligned} n^1_{\rm det}(r_{\rm a}, M_{\rm a}) &= \int {\rm d}r{\rm d}M\, \mathcal{N}^1_\star(r_{\rm a}, M_{\rm a}; r, M) \cdot \Gamma_1(r, M) \cdot p_{\rm tra}(M) \\ &= N^0_\star(\delta_{\rm obs}, L(M_{\rm a})) \cdot p_{\rm tra}(M_{\rm a}) \cdot \frac{\rm BF}{1-{\rm BF}} \int \frac{{\rm d}q}{\mathcal{A}^4}\, f(q)\, \Gamma_1\left(\frac{r_{\rm a}}{\mathcal{A}}, M_{\rm a}\right). \end{aligned} \tag{A18}$$

Finally,

$$\begin{aligned} n^2_{\rm det}(r_{\rm a}, M_{\rm a}) &= \int {\rm d}r{\rm d}M\, \mathcal{N}^2_\star(r_{\rm a}, M_{\rm a}; r, M) \cdot \Gamma_2(r, M) \cdot p_{\rm tra}(M) \\ &= N^0_\star(\delta_{\rm obs}, L(M_{\rm a})) \cdot \frac{\rm BF}{1-{\rm BF}} \int \frac{q{\rm d}q}{\mathcal{A}^3\mathcal{B}}\, f(q)\, \Gamma_2\left(\frac{r_{\rm a}}{\mathcal{B}}, qM_{\rm a}\right)\, p_{\rm tra}(qM_{\rm a}). \end{aligned} \tag{A19}$$

General formula for apparent occurrence rate—Using the above results, the apparent rate density,

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{N_\star(r_{\rm a}, M_{\rm a}) p_{\rm tra}(r_{\rm a}, M_{\rm a})} \times \sum_i n^i_{\rm det}(r_{\rm a}, M_{\rm a}), \tag{A20}$$

evaluates to

$$\begin{aligned} \Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{1+\mu({\rm BF}, M_{\rm a})} \times \Bigg\{ \Gamma_0(r_{\rm a}, M_{\rm a}) + \frac{\rm BF}{1-{\rm BF}} \Bigg[\int \frac{{\rm d}q}{\mathcal{A}^4}\, f(q)\, \Gamma_1\left(\frac{r_{\rm a}}{\mathcal{A}}, M_{\rm a}\right) \\ + \int \frac{q{\rm d}q}{\mathcal{A}^3\mathcal{B}}\, f(q)\, \Gamma_2\left(\frac{r_{\rm a}}{\mathcal{B}}, qM_{\rm a}\right) \frac{R(qM_{\rm a})}{R(M_{\rm a})} q^{-1/3} \Bigg] \Bigg\}. \end{aligned} \tag{A21}$$

This equation is used to derive Eqs. 24 and 28, and is numerically integrated to create Figs. 4, 6, and 7. It is validated in limits in which it is possible to write down the answer (e.g., Equation 11), and also against a Monte Carlo realization of the twin binary models (Secs. 3.1 and 5.1).

Noting the definition of μ in Equation 19, Equation A21 can also be expressed as

$$\begin{aligned} \Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{1+\mu({\rm BF}, M_{\rm a})} \cdot \Bigg[\Gamma_0(r_{\rm a}, M_{\rm a}) \\ + \mu({\rm BF}, M_{\rm a}) \cdot \left\langle \frac{1}{\mathcal{A}} \cdot \Gamma_1\left(\frac{r_{\rm a}}{\mathcal{A}}, M_{\rm a}\right) + \frac{q}{\mathcal{B}} \cdot \Gamma_2\left(\frac{r_{\rm a}}{\mathcal{B}}, qM_{\rm a}\right) \cdot \frac{R(qM_{\rm a})}{R(M_{\rm a})} q^{-1/3} \right\rangle \Bigg], \end{aligned} \tag{A22}$$

where the angle brackets denote averaging over $f(q)/\mathcal{A}^3$. This form shows that the fraction μ and mass ratio distribution $f(q)/\mathcal{A}^3$ of binaries *in the searchable volume* are enough to describe the contributions from binaries to Γ_{a}. The fraction μ gives the relative contributions from singles and binaries, and $f(q)/\mathcal{A}^3$ weights the contributions from binaries with various mass ratios. The apparent rate density at each mass ratio (*i.e.*, the terms in the angle brackets) is the rate density of systems that give the apparent $(r_{\mathrm{a}}, M_{\mathrm{a}})$, corrected for the overestimated transit probability around secondaries.

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; $r_p = 8 - 32 r_\oplus$; solar subset[a]
–	5±1	solar subset extended to $Kp < 16$
–	7.6±1.3	solar subset extended to $r_p > 5.6 r_\oplus$.
Moutou et al. (2013)	10±3	*CoRoT* average; $P \lesssim 10$ day, $r_p > 4 r_\oplus$
Petigura et al. (2017b)	$5.7^{+1.4}_{-1.2}$	$r_p = 8 - 24 r_\oplus$; $P = 1 - 10$ 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

Akeson, R. L., Chen, X., Ciardi, D., et al. 2013, Publications of the Astronomical Society of the Pacific, 125, 989

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

Bastien, F. A., Stassun, K. G., Pepper, J., et al. 2014, The Astronomical Journal, 147, 29

Batalha, N. M., Borucki, W. J., Koch, D. G., et al. 2010, The Astrophysical Journal Letters, 713, L109

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
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2016, The Astrophysical Journal, 828, 99
Ciardi, D. R., Beichman, C. A., Horch, E. P., & Howell, S. B. 2015, The Astrophysical Journal, 805, 16
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., & Howell, S. B. 2017, arXiv:1707.01942 [astro-ph]
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
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
Jenkins, J. M., Caldwell, D. A., & Borucki, W. J. 2002, The Astrophysical Journal, 564, 495
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

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., & Gaudi, B. S. 2005, The Astrophysical Journal, 631, 581

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

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

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

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

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

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]