

# *Annual Review of Astronomy and Astrophysics*

## Exoplanet Statistics and Theoretical Implications

Wei Zhu<sup>1,2</sup> and Subo Dong<sup>3,\*</sup>

<sup>1</sup>Department of Astronomy, Tsinghua University, Beijing 100084, China;  
email: weizhu@tsinghua.edu.cn

<sup>2</sup>Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto,  
Ontario M5S 3H8, Canada

<sup>3</sup>Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China;  
email: dongsubo@pku.edu.cn

**ANNUAL  
REVIEWS CONNECT**

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Astron. Astrophys. 2021. 59:291–336

First published as a Review in Advance on  
June 11, 2021

The *Annual Review of Astronomy and Astrophysics* is  
online at [astro.annualreviews.org](http://astro.annualreviews.org)

<https://doi.org/10.1146/annurev-astro-112420-020055>

Copyright © 2021 by Annual Reviews.  
All rights reserved

\*Corresponding author

### Keywords

planetary systems, orbital properties, planet formation, dynamical evolution

### Abstract

In the past few years, significant advances have been made in understanding the distributions of exoplanet populations and the architecture of planetary systems. We review the recent progress of planet statistics, with a focus on the inner  $\lesssim 1$ -AU region of planetary systems that has been fairly thoroughly surveyed by the *Kepler* mission. We also discuss the theoretical implications of these statistical results for planet formation and dynamical evolution.

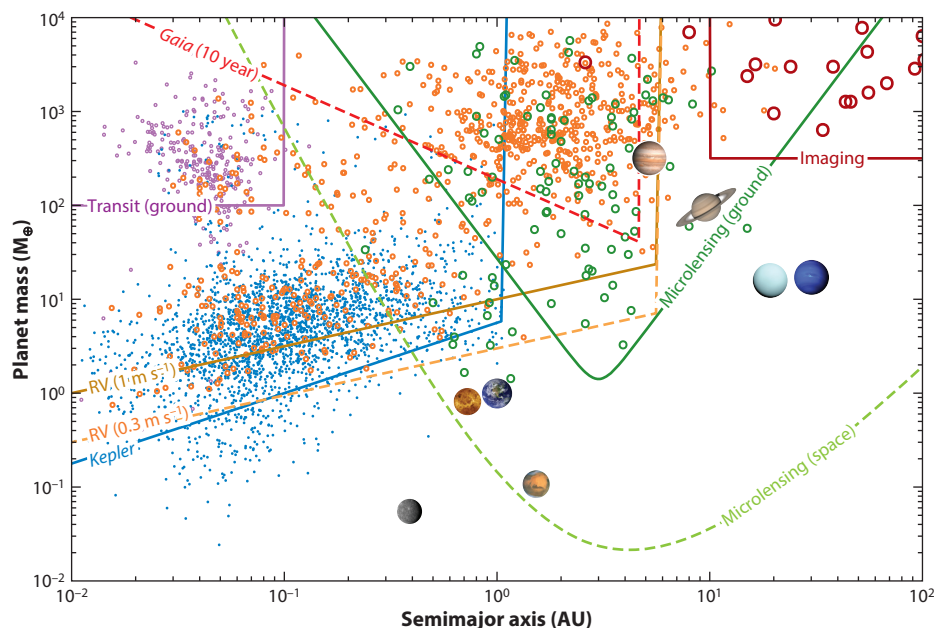
## Contents

1. INTRODUCTION .....	292
1.1. On Defining and Interpreting Planet Occurrence Rate .....	294
1.2. On Inferring the Frequency of Planets .....	295
2. THE INNER PLANETARY SYSTEM .....	297
2.1. Planet Distribution in the Period–Radius Plane .....	298
2.2. Mutual Inclinations and the Intrinsic Multiplicity .....	305
2.3. Eccentricity Distribution .....	309
2.4. Intrasytem Variation .....	310
2.5. Dependence of Planet Statistics on Stellar Properties .....	314
3. THE OUTER PLANET POPULATION .....	318
3.1. Planet Frequency .....	318
3.2. The Inner–Outer Correlation .....	318
3.3. Mass Ratio Function from Microlensing .....	321
3.4. Free-Floating Planets .....	323
4. THEORETICAL IMPLICATIONS .....	325
4.1. A Brief Overview of Theories .....	325
4.2. Constraints from Observations .....	326
5. SUMMARY AND DISCUSSION .....	327

## 1. INTRODUCTION

“Who ordered that?” said the theorist I. Rabi when learning about the unexpected discovery of muons in 1936. Little did particle physicists know that it would only be the beginning of uncovering a puzzling particle zoo filled with diverse particles in the next three decades, until revolutionary theoretical insights were developed to classify the elementary particles. Now nearly three decades since the astonishing discovery of a hot Jupiter (Mayor & Queloz 1995), the exoplanet zoo is ever growing—whenever the detection territories grow in breadth or depth, nature appears to be teeming with new species. Theorists working on planet formation and evolution face distinctly different sets of challenges from particle physicists: In the popular paradigm, forming planets from dust grains is a daunting march spanning tens of orders of magnitudes in mass and involves many physical processes that are too complex for first-principle calculations. In hindsight, it should probably be of little surprise that a theory involving such complicated physics, which was anchored by the sole sample of our Solar System, would have limited predictive success.

We review the recent progress of planet statistics and identify patterns emerging from the known thousands of exoplanets that cover a broad region of the parameter space (see **Figure 1**). Robustly identifying patterns in the intrinsic distributions of planets can stimulate and test theories. Conversely, theoretical advances may also beam the searchlight on fresh observational ground, as exemplified by the development of the photoevaporation theory leading to the recent discovery of a radius valley (see Section 2.1.4). Since the last *Annual Reviews* article on exoplanet populations (Winn & Fabrycky 2015), the field of planet statistics has made significant progress. In particular, the large and homogeneous planet sample from the NASA *Kepler* mission (Borucki et al. 2010) has provided the best source for statistical studies, but a major shortcoming of the *Kepler* data was the initial lack of accurate stellar parameters for both the planet hosts



**Figure 1**

Mass versus semimajor axis of known planets, based on the “Confirmed Planets” list from the NASA Exoplanet Archive (Akeson et al. 2013; acquired in September of 2020) and the reliable *Kepler* planet candidates (see Section 2 for more details). Using different colors, we differentiate planet detections as well as the approximate sensitivity curves from ground-based transit (purple), *Kepler* survey (blue), RV (orange and brown), microlensing (green), and direct imaging (red). The masses of the *Kepler* detections are estimated from the measured radii according to the Chen & Kipping (2017) mass–radius relation. The sensitivity curve of *Kepler* is also converted in a similar way from that measured on the period–radius plane (see Section 2.1). The sensitivity curve for the 10-year *Gaia* astrometry survey is also shown in red, for which we assume a Sun-like host at 20 pc and require a 3- $\sigma$  detection over the expected precision. For space-based microlensing, we adopt the sensitivity curve of the microlensing survey that will be performed by the *Nancy Grace Roman Space Telescope* (formerly known as WFIRST or *Wide Field Infrared Survey Telescope*; Penny et al. 2019). The Solar System planet images are shown at their corresponding locations. Data for the dashed green line is taken from Penny et al. (2019). Abbreviation: RV, radial velocity.

and the target stars (i.e., the parent sample). In the past few years, substantial efforts have been dedicated to systematically characterize the *Kepler* sample and, thus, unleash its potential for statistical studies. These include asteroseismology (e.g., Chaplin & Miglio 2013, Van Eylen & Albrecht 2015), the *Gaia* data releases (Gaia Collab. et al. 2016, 2018), follow-up spectroscopic programs such as the LAMOST (Large Sky Area Multi-Object Fibre Spectroscopic Telescope; Cui et al. 2012, Zhao et al. 2012)–*Kepler* survey (e.g., Dong et al. 2014b, De Cat et al. 2015, Zong et al. 2018) and the California–*Kepler* Survey (CKS; Johnson et al. 2017, Petigura et al. 2017), as well as many projects of the *Kepler* Follow-up Observation Program (KFOP; Furlan et al. 2017). Furthermore, substantial works to understand the *Kepler* pipeline detection efficiency and vetting false positives have much improved the reliability of *Kepler* statistical inference (e.g., Christiansen et al. 2015, Morton et al. 2016). Last but not least, in-depth developments have been recently made to disentangle the intricate observational biases of multiplanet systems. These efforts have made it possible to offer new insights into planet distributions and architectures.

In this review, we first clarify in Sections 1.1 and 1.2 several common confusions in exoplanet statistical studies. Then, we discuss planet distributions in the inner ( $\lesssim 1$  AU) and the outer ( $\sim 1\text{--}10$  AU) regions in Sections 2 and 3, respectively. The former is focused on results from the *Kepler* mission, and the latter includes updated results from radial velocity (RV) and gravitational microlensing. A brief discussion of the free-floating planets (FFPs) from microlensing is also provided. We focus on planets around  $\gtrsim 1$ -Gyr-old stars, whereas planets orbiting young stars found by direct imaging are not discussed (see the review by Bowler 2016). The implications to theories of planet formation and evolution are discussed in Section 4. Finally, in Section 5, we summarize and outline the promising directions for future developments.

## 1.1. On Defining and Interpreting Planet Occurrence Rate

Many statistical studies focus on deriving the intrinsic occurrence rate (or the often interchangeably used term, frequency) of planets. But from one study to another, the same term can carry different meanings. In the following, we clarify these different definitions to avoid further misinterpretations.

In most studies, the derived occurrence rate is the average number of planets per star, and we denote it as  $\bar{n}_p$ , which is defined as

$$\bar{n}_p \equiv \frac{\text{Total \# of planets}}{\text{Total \# of stars}}. \quad 1.$$

Here, a planet is restricted to lie within a predefined parameter space, often in the period–radius plane (for the transit method) or the period–mass (or minimum mass  $m_p \sin i$ ) plane (for the RV method). Similarly, a star is restricted to a star-like target of predefined properties. Because a large fraction of such stars may actually have unresolved stellar companions, the correction for the impact of the stellar binarity can be important for the inference of the planet-formation efficiency (see Section 2.5.1).

Another important quantity sometimes referred to as occurrence rate is the fraction of stars with planets,  $F_p$ :

$$F_p \equiv \frac{\text{Total \# of planetary systems}}{\text{Total \# of stars}}. \quad 2.$$

Here, a planetary system has at least one planet existing in a predefined parameter space. By definition,  $F_p \leq 1$ , so it is usually reported as a percentage. However, an occurrence rate reported as a percentage (i.e., “X% of stars have planets”) does not necessarily mean that it is the fraction of stars that are hosts of planets, because  $\bar{n}_p$  is also frequently reported as a percentage.

To distinguish between the two definitions, we refer to  $\bar{n}_p$  as the frequency of planets and  $F_p$  as the frequency of planetary systems. The ratio of the two measures the average number of planets per planetary system (within a predefined parameter space), which we call average planet multiplicity and denote as  $\bar{m}_p$ :

$$\bar{m}_p \equiv \frac{\bar{n}_p}{F_p} = \frac{\text{Total \# of planets}}{\text{Total \# of planetary systems}}. \quad 3.$$

*Kepler* data suggest that multiplanet systems are common, so usually  $\bar{m}_p$  is larger than unity, and consequently  $\bar{n}_p$  and  $F_p$  substantially differ from each other. They only become similar when the average planet multiplicity  $\bar{m}_p \rightarrow 1$ , which can happen when either (a) a category of planets with low intrinsic multiplicity (e.g., short-period giant planets) is concerned or (b) the parameter space of interest is small enough that systems with more than one such planet are rare.

The three quantities,  $\bar{n}_p$ ,  $F_p$ , and  $\bar{m}_p$ , are all important for testing theories. To provide a simple example, with only  $\bar{n}_p$  measured to be unity, it is possible that all stars have one planet ( $F_p = 100\%$  and  $\bar{m}_p = 1$ ) or that half of the stars have two planets ( $F_p = 50\%$  and  $\bar{m}_p = 2$ ). These two cases obviously demand different theoretical explanations.

Observationally, the derivations of  $\bar{n}_p$  and  $F_p$  have rather different requirements and follow different procedures. It is generally more straightforward to derive  $\bar{n}_p$ , because correcting the detectability of individual planets concerns observables directly measurable from surveys (e.g., orbital periods and planet sizes for transit, assuming that the properties of the stars are known). In contrast, the detectability of a planetary system usually concerns the intrinsic architecture of the system, including the planet multiplicity and distributions of the orbital and physical parameters, many of which may not be directly observable, so the derivation of  $F_p$  can rely on assumptions of these unknowns. This is especially an issue in transit surveys: The derivation of  $F_p$  requires assumptions about the mutual inclinations between planets, and different assumptions can lead to fairly different values of  $F_p$  (see Section 2.2).

In deriving the two frequencies, statistical studies involving multiplanet systems usually treat the planet occurrence as a Poisson process. This may be a reasonable assumption in the derivation of the planet frequency  $\bar{n}_p$ , but it can lead to unreliable results in the derivation of the planetary system frequency  $F_p$ . This Poisson process assumption implies that the presences of individual planets in the same system are independent and that their physical and orbital properties are independent of the properties of other planets or of the host star. As discussed later in this review, such an assumption breaks down in certain circumstances. Below, we provide a specific example to demonstrate its impact on the planetary system frequency. The fractions of Sun-like stars with cold giant planets and with planets that *Kepler* is sensitive to are 10% and 30%, respectively. The fraction of such stars with at least one planet in the joint parameter space would be  $1 - (1 - 30\%) \times (1 - 10\%) = 37\%$  under the Poisson process assumption. However, this frequency is determined to be  $\sim 30\%$  as a result of the strong correlation between the inner and the outer planets (Section 3.2). The correlations (or sometimes anticorrelations) between the occurrences of planets around the same host also suggest that one may not be able to extrapolate a parameterized distribution of the planetary system frequency to a parameter space that is not covered by the data.

A number of studies have reported  $F_p$  by using the detectability of the first detected (or the most detectable) planet in the system as that of the whole system (e.g., Cumming et al. 2008, Mayor et al. 2011, Fressin et al. 2013, Petigura et al. 2013). This approach does not require assumptions on planet multiplicity or architecture. However, as the detectability of any planet is no greater than the detectability of the system it resides in, this approach typically tends to overestimate  $F_p$  (Zhu et al. 2018b).

## 1.2. On Inferring the Frequency of Planets

In this section, we discuss the commonly used methods of inferring the frequency of planets  $\bar{n}_p$  from a statistical survey of  $N_\star$  target stars. A popular method is the so-called inverse detection efficiency method (IDEM), which has been used extensively in the literature, including in many influential studies (e.g., Mayor et al. 2011; Howard et al. 2012; Fressin et al. 2013; Petigura et al. 2013; Dressing & Charbonneau 2013, 2015). For our illustrative survey, the average number of planets per star according to IDEM is

$$\bar{n}_p^{\text{IDEM}} = \frac{1}{N_\star} \sum_{i=1}^{N_p} \frac{1}{p_i} = \frac{N_p}{N_\star} \left\langle \frac{1}{p} \right\rangle. \quad 4.$$

Here,  $p_i$  is the survey detection efficiency of the  $i$ th of  $N_p$  detected planets and  $\langle \cdot \rangle$  is the average over all detected planets. IDEM is intuitive, simple to perform, and computationally efficient, as it does not require computing the detection efficiencies of null detections (which are usually the majority of the targets), so it is useful in getting a rough estimate of the underlying frequency. However, this method is not rigorously established in the probability theory and can potentially lead to biased results (Foreman-Mackey et al. 2014, Hsu et al. 2018). Specifically, with a low detection efficiency and a small number of detections, Hsu et al. (2018) found that IDEM often leads to underestimated  $\bar{n}_p$  because the actual detections typically come from targets with larger-than-average sensitivities. IDEM can also suffer substantial fluctuations because of the inversion of the (typically small) detection efficiency.

An approach with sound statistical basis is modeling planet occurrence as a Poisson process and performing maximum likelihood (ML) analysis (e.g., Tabachnik & Tremaine 2002, Cumming et al. 2008, Gould et al. 2010, Youdin 2011, Dong & Zhu 2013, Burke et al. 2015). In a given bin that has  $N_p$  planet detections, Youdin (2011, see their section 3.1) and Foreman-Mackey et al. (2014, see their appendix A) show that the ML estimator for planet frequency is

$$\bar{n}_p^{\text{ML}} = \frac{N_p}{\sum_{j=1}^{N_*} p_j} = \frac{N_p}{N_*} \frac{1}{\langle p \rangle} = \frac{N_p}{N_*^{\text{eff}}}; \quad N_*^{\text{eff}} \equiv N_* \langle p \rangle. \quad 5.$$

Here,  $p_j$  is the planetary detection efficiency in the bin for the  $j$ th star, regardless of whether the star yields any actual planet detection or not, and  $N_*^{\text{eff}}$  is the effective number of target stars. Unlike in Equation 4, the average here is performed among all target stars. Compared to IDEM, this method is computationally more expensive while being statistically superior. It is more robust against fluctuations in the efficiencies of individual detections (as well as null detections) because the averaging is performed on  $p$  rather than  $1/p$ .

Next, we elaborate on incorporating the above approach into the Bayesian framework following the simplified Bayesian model of Hsu et al. (2018, see their appendix B) but with some corrections. The posterior probability distribution of planet frequency  $\bar{n}_p$  for the statistical sample is given by

$$P(\bar{n}_p | N_p, N_*^{\text{eff}}) \propto P(N_p | \bar{n}_p, N_*^{\text{eff}}) P_{\text{pri}}(\bar{n}_p). \quad 6.$$

The first term on the right-hand side quantifies the probability (or likelihood) of having the  $N_p$  detections for a given rate  $\bar{n}_p$ , which under the Poisson process assumption is described by a Gamma distribution.<sup>1</sup> The second term,  $P_{\text{pri}}(\bar{n}_p)$ , is the prior distribution of  $\bar{n}_p$ . If a conjugate prior is assigned as a Gamma distribution with a shape parameter  $\alpha_0$  and a rate parameter  $\beta_0$ , the resulting posterior distribution is then a Gamma distribution with the shape parameter  $\alpha_0 + N_p$  and the rate parameter  $\beta_0 + N_*^{\text{eff}}$ :

$$P(\bar{n}_p | N_p, N_*^{\text{eff}}) \propto \bar{n}_p^{\alpha_0 + N_p - 1} e^{-\bar{n}_p (\beta_0 + N_*^{\text{eff}})}. \quad 7.$$

For a flat prior on  $\bar{n}_p$ , the two parameters are  $\alpha_0 = 1$  and  $\beta_0 = 0$ , respectively. For completeness, the mean and standard deviation of this Gamma distribution posterior are

$$\mu(\bar{n}_p) = \frac{\alpha_0 + N_p}{\beta_0 + N_*^{\text{eff}}}, \quad \sigma(\bar{n}_p) = \frac{\sqrt{\alpha_0 + N_p}}{\beta_0 + N_*^{\text{eff}}}. \quad 8.$$

<sup>1</sup>A Gamma distribution can be parameterized in terms of a shape parameter  $\alpha$  ( $>0$ ) and a rate parameter  $\beta$  ( $>0$ ). The probability density function of a variable  $x$  is  $f(x; \alpha, \beta) = (\beta^\alpha x^{\alpha-1} e^{-\beta x}) / \Gamma(\alpha) \propto x^{\alpha-1} e^{-\beta x}$ , where  $\Gamma(\alpha)$  is the Gamma function evaluated at  $\alpha$ .

The first expression reduces to the ML estimator of Equation 5 if a log-flat prior on the planet frequency  $\bar{n}_p$  is assumed (i.e.,  $\alpha_0 = \beta_0 = 0$ ). The expressions given by Equation 8 provide easy-to-use estimates to report when the number of detections is relatively large. However, when small or null detections are involved, the posterior probability distribution is fairly non-Gaussian. It is then more appropriate to report the median value, the 68% credible interval, and/or the 95% upper limit, all of which can be derived from the cumulative posterior probability distribution. It is also worth noting that, in the case of null detections, a meaningful upper limit on  $\bar{n}_p$  cannot be derived with the log-flat prior because the shape parameter becomes zero and the Gamma distribution is undefined. We show in Section 2.1 an application of the Bayesian approach to derive the frequency of planets in the *Kepler* parameter space.

## 2. THE INNER PLANETARY SYSTEM

We review in this section planet statistics in the inner region ( $\lesssim 1$  AU of Sun-like stars), which is well explored thanks to thousands of planets detected by the RV and transit techniques. We focus on the best statistical probe by far of the inner region—the large and uniform sample from the *Kepler* mission, which is sensitive to transiting planets with radii  $R_p$  down to  $\sim R_\oplus$  and orbital periods  $P$  up to  $\sim 1$  year (Borucki et al. 2010).

We first derive a clean baseline sample based on the final *Kepler* data release (DR25; Thompson et al. 2018) and the improved stellar parameters from Berger et al. (2020b). The latter work combines the astrometric measurements from *Gaia* DR2 (Gaia Collab. et al. 2018) with the available photometric and spectroscopic information to yield stellar radii with a median uncertainty of 4%. Starting from the DR25 planet catalog, we have removed planet candidates with (a) transit signal-to-noise ratio (S/N) below the nominal threshold (S/N = 7.1), (b) NASA Exoplanet Archive ([exoplanetarchive.ipac.caltech.edu](http://exoplanetarchive.ipac.caltech.edu)) disposition flag being false positive, (c) the derived planetary radius  $R_p > 20 R_\oplus$ , (d) the orbital period  $P > 400$  days, and (e) the best-fit transit impact parameter  $b > 1$ . We restrict our discussion to Sun-like stars that are defined as main-sequence stars (as classified by Berger et al. 2018) with effective temperatures between 4,700 K and 6,500 K. The bulk of this section is about planets around Sun-like hosts, and topics such as correlations with various stellar properties, such as stellar mass, metallicity and binarity, are discussed in Section 2.5.

The baseline sample contains 2,525 planet detections around 98,213 Sun-like stars. Of all the transiting planets, 1,451 are found in systems with only one detected transiting planet<sup>2</sup> and the remaining 1,074 are from systems with multiple detected transiting planets. The average observed multiplicity rate, namely the average fraction of planets from known multiplanet systems, is 42.5%. This is a lower limit on the intrinsic multiplicity rate, as many of the single-planet systems seen in *Kepler* are likely part of intrinsic multiplanet systems (see details in Section 2.2). The observed transit multiplicity distribution in the sample is

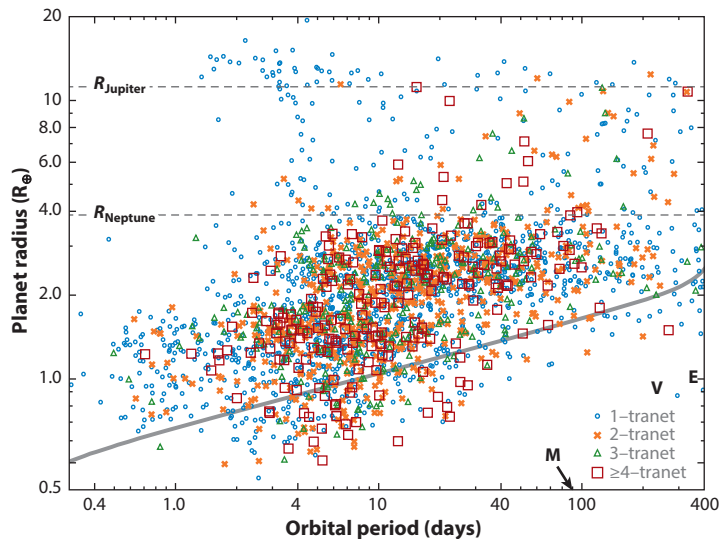
$$(N_1, N_2, N_3, N_4, N_5, N_6, N_7) = (1,451, 278, 97, 37, 12, 2, 1), \quad 9.$$

and no system has more than seven transiting planets.<sup>3</sup> **Figure 2** illustrates the planets in our sample in the period–radius plane. Different multiplicities of transiting planets are shown with different symbols.

<sup>2</sup>We sometimes use the contraction “tranet” to stand for “transiting planet” in the text and figure legends and captions.

<sup>3</sup>Note that the only system in our sample with seven transiting planets, Kepler-90, has been found to contain one additional planet candidate (Shallue & Vanderburg 2018). However, this additional candidate was not found by the *Kepler* DR25 pipeline and thus is not included.





**Figure 2**

The close-in ( $P < 400$  days) *Kepler* planets in the period–radius plane, with various symbols and colors indicating the observed multiplicity (note that we use the contraction “tranet” to refer to “transiting planet” in the legend). The gray solid curve indicates the median detection efficiency of the planet search pipeline. The Solar System planets in the inner region, namely Mercury, Venus, and Earth, are denoted with their first letters. The median precision on the planetary radius is  $\sim 7\%$ . The radii of Jupiter and Neptune are shown with horizontal dashed lines. The radius valley at  $\sim 2 R_{\oplus}$  (see Section 2.1.4) is visible.

## 2.1. Planet Distribution in the Period–Radius Plane

With the above statistical sample, we derive the planet frequencies in the Bayesian framework of Section 1.2. The parameter space in the period–radius plane is divided into logarithmically equally spaced cells.<sup>4</sup> In each cell, the number of planet detections,  $N_p$ , is found and the average detection efficiency,  $\langle p \rangle$ , is computed via

$$\langle p \rangle = \frac{\int_{R_{p,\min}}^{R_{p,\max}} \int_{P_{\min}}^{P_{\max}} (R_{\odot}/a) S(P, R_p) d \ln P d \ln R_p}{\int_{R_{p,\min}}^{R_{p,\max}} \int_{P_{\min}}^{P_{\max}} d \ln P d \ln R_p}. \quad 10.$$

Here  $R_{p,\min}$ ,  $R_{p,\max}$ ,  $P_{\min}$ , and  $P_{\max}$  denote the boundaries of the cell, and  $R_{\odot}/a$  is approximately the transit geometric probability at semimajor axis  $a$  around a Sun-like host. The sensitivity due to survey detection thresholds at a given period and radius,  $S(P, R_p)$ , is computed with the *Kepler*PORTs code (publicly available at <https://github.com/nasa/KeplerPORTs>), which was first developed by Burke et al. (2015) and further updated for *Kepler* DR25 (Burke & Catanzarite 2017a) by incorporating results of transit injection and recovery tests for the final *Kepler* pipeline (Burke & Catanzarite 2017b, Christiansen et al. 2020). Updated stellar parameters were used to derive the mean sensitivity curve.

We adopt a flat prior on  $\bar{n}_p$ , and its posterior distribution is then described by the Gamma distribution of Equation 7 with  $\alpha_0 = 1$  and  $\beta_0 = 0$ . For cells with  $\leq 2$  detections, we report the

<sup>4</sup>Because the typical precisions of planetary period and radius are much smaller than the cell sizes, we ignore the uncertainties of planetary parameters. See Foreman-Mackey et al. (2014) for how to incorporate the planet parameter errors in the analysis.





**Figure 3**

This figure illustrates the planet frequencies ( $\bar{n}_p$ ) and the observed multiplicity fractions based on the planet sample in **Figure 2**. The numbers with error bars are the percentages of planets with periods and radii within the given cell. If there are less than three detections found within the cell, then the 95% upper limit is reported instead. These upper limits are highlighted in red. The fraction in each cell denotes the observed multiplicity fraction, namely the fraction of planets in that cell found to reside in multiplanet systems. We use “N/A” for cells with less than three detections. The red dashed lines mark the regions corresponding to hot Jupiters, hot Neptune “desert,” and ultrashort-period planets.

95% upper limits, whereas for the rest the means and the standard deviations given by Equation 8 are reported as the measurements and associated uncertainties, respectively. We have verified that the deviation between the mean and the median is substantially smaller than the uncertainty for all relevant cells.

The derived planet frequency map is shown in **Figure 3**. For any cell with more than two detections, we also indicate the observed multiplicity rate of planets in the cell. Again, these multiplicity rates represent the lower limits on the fraction of planets in those cells that reside in multiplanet systems. We summarize several key results below.

- The integrated planet frequency is  $\bar{n}_p = 1.23 \pm 0.06$  for planets with radii in the range of 1–20  $R_{\oplus}$  and orbital periods of up to 400 days. This is broadly consistent with results from previous studies (e.g., Fressin et al. 2013, Petigura et al. 2018, Hsu et al. 2019). As stressed in Section 1.1, statistical analyses like this one do not yield the fraction of stars with planets  $F_p$ , as the impacts of the multiplicity and the mutual inclination have not been taken into account (see Section 2.2).
- As has been clear since the earliest *Kepler* statistical studies, there are generally many more small planets with radii  $R_p \lesssim 4 R_{\oplus}$  than larger ones for orbital periods  $P < 400$  days. Planet

frequencies tend to increase from the upper left (large  $R_p$  and small  $P$ ) toward the lower right (small  $R_p$  and large  $P$ ). In other words, the intrinsic radius distribution is dependent on the orbital period (e.g., Dong & Zhu 2013, Foreman-Mackey et al. 2014, Hsu et al. 2018). There exist some local regions where the general trends break down, such as the radius valley (see Section 2.1.4).

- Sub-Earths ( $R_p < 1 R_\oplus$ ) and Earth-sized planets in Earth-like orbits are not well probed by *Kepler*, and thus estimates of their frequencies are most susceptible to the uncertainties of survey sensitivity estimates. As a result, there remain large discrepancies on their intrinsic frequencies in the literature (see Winn & Fabrycky 2015, their table 2, and Burke et al. 2015, their figure 17).
- *Kepler* planets commonly reside in multiplanet systems in most parts of the period–radius plane, with some notable exceptions such as the hot Jupiter region (Steffen et al. 2012; see Section 2.1.1 for more discussion). The intrinsic multiplicity rates are likely higher than the observed multiplicity rates shown in **Figure 3**. We defer to Section 2.2 for further discussion.

The above method to derive the planet frequency  $\bar{n}_p$  is nonparametric. An alternative approach employs a parameterized planet distribution function and then constrains the associated parameters. The parametric approach has been widely used in statistical studies of various detection techniques, including transit (e.g., Youdin 2011, Howard et al. 2012, Dong & Zhu 2013, Burke et al. 2015), RV (e.g., Tabachnik & Tremaine 2002, Cumming et al. 2008), and microlensing (e.g., Gould et al. 2010, Clanton & Gaudi 2016, Suzuki et al. 2016) studies. It is also commonly used in simulations of generating synthetic planetary systems (e.g., Mulders et al. 2018, He et al. 2019). The commonly adopted planet distribution function is separable between the orbital period (or semimajor axis) and the planetary radius (or mass):

$$\frac{d^2 N}{d \ln P d \ln R_p} \propto \frac{dN}{d \ln P} \frac{dN}{d \ln R_p}. \quad 11.$$

The distributions of the orbital period and the planetary radius are usually parameterized as power laws or broken power laws. The use of such a separable function implicitly assumes that the period (radius) distribution is independent of the planetary radius (period). As discussed above, such an assumption is not valid for the inner planetary system. It is likely not valid for planet distributions in other regions of the parameter space either. The implications of this failure on the derived frequencies from the parametric method and on the theoretical interpretations of the underlying population have not been fully explored. In what follows, we provide brief discussions about selected regions in the period–radius plane.

**2.1.1. Hot Jupiters.** As the first type of exoplanets found around solar-type stars (Mayor & Queloz 1995), hot Jupiters ( $8 R_\oplus < R_p < 20 R_\oplus$  and  $P < 10$  days) remain interesting and exciting targets for both observational and theoretical purposes. Here, we only review the frequency and multiplicity of hot Jupiters in the current context and refer interested readers to the recent review by Dawson & Johnson (2018) for more in-depth discussion about the hot Jupiter population.

There is a long-standing discrepancy between the hot Jupiter frequencies inferred from RV and transit surveys (e.g., Gould et al. 2006a, Wright et al. 2012; see Santerne et al. 2016, their table B9 for an incomplete list of references). For example, our statistical sample yields a rate of  $(0.62 \pm 0.09\%)$ , which is in good agreement with previous studies of the hot Jupiter frequency in the *Kepler* field (e.g., Howard et al. 2012, Fressin et al. 2013, Santerne et al. 2016), whereas the RV surveys of stars in the Solar Neighborhood report rates that are typically a factor of  $\sim 2$  higher (0.9–1.2%; Mayor et al. 2011, Wright et al. 2012). It was suggested that the discrepancy could be caused by the different stellar properties, such as age, metallicity, and binary fraction, between the

**RV and transit samples.** This has been tested by several follow-up studies of the *Kepler* sample. The *Kepler* stars are only slightly subsolar on average ( $\langle[\text{Fe}/\text{H}]\rangle_{\text{Kepler}} \approx -0.04$ ; Dong et al. 2014b), and their metallicity differences with the RV targets ( $\langle[\text{Fe}/\text{H}]\rangle_{\text{RV}} \approx 0.0$ ) seem to be too small to fully account for the discrepancy even given the steep dependence of hot Jupiter frequency on metallicity (Guo et al. 2017). The unresolved binaries are also unlikely to substantially change the hot Jupiter frequency in the *Kepler* sample (Bouma et al. 2018). However, because RV surveys preferentially exclude close ( $\sim 1\text{--}50$  AU) stellar binaries from their samples, this discrepancy in hot Jupiter frequencies between transit and RV surveys can potentially be resolved if the formation of hot Jupiters is suppressed in such close binary systems (Moe & Kratter 2019). Searching for stellar companions of transiting hot Jupiters (e.g., Ngo et al. 2016) and making comparisons with field stars constitute a promising way to further test this possibility.

As shown in **Figure 2**, 1 out of the 49 hot Jupiters in our statistical sample, Kepler-730b, has a nearby small planet companion (Zhu et al. 2018a, Cañas et al. 2019). As of this writing, only two other hot Jupiters, WASP-47b (Becker et al. 2015) and TOI-1130c (Huang et al. 2020), are known to share the same property. Our statistical sample suggests that  $\sim 2\%$  ( $< 9.7\%$ , 95% upper limit) of hot Jupiters have nearby ( $\lesssim 20$  days), small ( $\sim 1\text{--}4 R_{\oplus}$ ), and nearly coplanar companions (see also Steffen et al. 2012 for the constraint on noncoplanar companions). This low multiplicity rate of hot Jupiters supports the general idea that most of them have undergone some large-scale migrations to arrive at the current locations (e.g., Lin et al. 1996, Rasio & Ford 1996, Weidenschilling & Marzari 1996). We refer to Dawson & Johnson (2018) for more in-depth discussion on this topic.

**2.1.2. Hot Neptune desert.** The region located at  $P \lesssim 4$  days and  $2 R_{\oplus} \lesssim R_p \lesssim 8 R_{\oplus}$  lands in the so-called hot Neptune (or sub-Jovian) desert (e.g., Szabó & Kiss 2011, Beaugé & Nesvorný 2013, Mazeh et al. 2016, and references therein), which is considered underpopulated, especially when inspecting mixed planet samples found in surveys with different detection sensitivities (e.g., ground-based transits and *Kepler*). This desert is however not that barren: The total planet frequency enclosed in the above region is  $0.61 \pm 0.07\%$  from our statistical analysis (see **Figure 3**), making this hot Neptune desert similarly populated as the hot Jupiter region (see also Dong et al. 2018). Although the above frequency is derived for a rectangular region in the period–radius plane, it is worth noting that the boundaries of this desert region are better described as a triangle and extend out to 5–10 days in  $m_p$  versus  $a$  and  $R_p$  versus  $P$  planes (see Mazeh et al. 2016, their figures 1 and 4). Dong et al. (2018) found that the frequency of planets inside this region depends on the host star metallicity in a way similar to the frequency of hot Jupiters, and they dubbed this population as Hoptunes (rather than hot Neptunes) to reflect that not all of them were known to be Neptune-like physically. Out of our baseline sample of 61 planets in this region, 14 are observed to have planetary companions, and the periods for the majority of these companions are within 10 days. The observed multiplicity rate is thus 23%, which is lower than the *Kepler* average while higher than that of hot Jupiters (see also Dong et al. 2018). We refer to Dawson & Johnson (2018) for more discussion on the connection of this population with close-in Jupiters and related theoretical implications.

A number of theories have been proposed to explain the formation of planets in this region (e.g., Kurokawa & Nakamoto 2014, Lundkvist et al. 2016, Matsakos & Königl 2016, Bailey & Batygin 2018, Owen & Lai 2018). The leading explanations of its triangular boundaries invoke photoevaporation (see more discussion in Section 2.1.4) and tidal effects following the high-eccentricity migration. The upper boundary is best explained as the tidal disruption barrier for gas giants following their high-eccentricity migrations (Matsakos & Königl 2016, Owen & Lai 2018). More massive planets can be tidally circularized closer to the star without tidal disruption, resulting in the negative slope of the upper boundary. It has been proposed that the

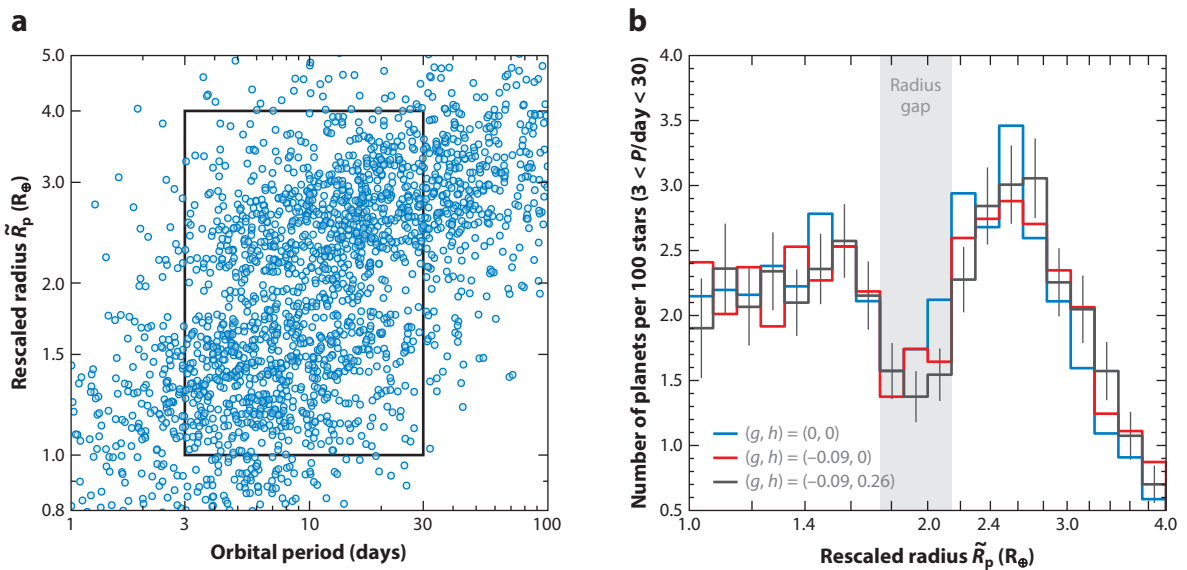
same mechanism also produces the lower boundary, with the positive slope resulting from a mass–radius relation of small planets that is different from the relation of giant planets (Matsakos & Königl 2016). However, this mechanism may not be able to explain the planets that are in or near the desert region and reside in multiplanet systems. An alternative theory, proposed by Owen & Lai (2018), suggests that the lower boundary is better explained by the photoevaporation of highly irradiated planets and that the positive slope results from the fact that the photoevaporation mechanism is more effective if the planet is closer to the host star. There has been a growing interest for planets in this region with the TESS (*Transiting Exoplanet Survey Satellite*; e.g., Armstrong et al. 2020, Burt et al. 2020) mission, and follow-up studies of such planets will soon allow for a better understanding of their physical properties and formation mechanisms.

**2.1.3. Ultra-short-period planets.** Planets with radii of  $0.5\text{--}2 R_{\oplus}$  and periods of  $P \lesssim 1$  day known as ultrashort-period planets (USPs) represent a rather extreme planet population. The period threshold for USPs at one day corresponds to an equilibrium temperature of  $\sim 2,000$  K for a Sun-like host, which is hot enough to sublimate dust grains. Below, we briefly summarize several key properties of USPs and refer interested readers to the recent comprehensive review by Winn et al. (2018) for more discussion about this extreme planet population.

Our statistical analysis yields  $\bar{n}_p = (0.39 \pm 0.04)\%$  for USPs. This is in general agreement with the result of Sanchis-Ojeda et al. (2014), whose specialized pipeline yields  $\bar{n}_p = (0.51 \pm 0.07)\%$  for planets with radii in the range of  $0.8\text{--}2 R_{\oplus}$  and  $P < 1$  day. Out of the 81 USPs in our sample, 16 are found with outer planetary companions, indicating an observed multiplicity rate of 20%. The true multiplicity rate is probably much higher, because USPs can be largely misaligned relative to the outer planetary companions (Dai et al. 2018, Petrovich et al. 2019). In 13 of the 16 multiplanet systems involving USPs, the closest outer companion has  $P_c \lesssim 10$  days, and the USP is usually farther apart in terms of the period ratio from the rest of the planets in the same system (see also Steffen & Farr 2013).

The highly irradiative environment at subday orbit implies that USPs are unlikely to have formed in situ. Partially because of the comparable rates between hot Jupiters and USPs, it had been suggested that USPs could be the surviving cores of tidally disrupted hot Jupiters (Jackson et al. 2013), but this was not supported by several pieces of evidence including the lack of strong host metallicity dependence (Winn et al. 2017) and the relatively high multiplicity rate compared to hot Jupiters. A more plausible scenario is that the USPs have arrived at their current locations without losing much of their initial mass. One way of achieving this is the gradual decay of the orbit due to the tidal dissipation within the host star (Lee & Chiang 2017). Alternatively, the proto-USP planet may have been sent to an eccentric (and misaligned) orbit following the dynamical interactions with other planets in the system, and then the orbit may have decayed and circularized due to the tidal dissipation within the planet (Schlaufman et al. 2010, Petrovich et al. 2019, Pu & Lai 2019). This latter model sees its support in the relatively large mutual inclinations of USPs (Dai et al. 2018). Additionally, in order for the tidal inspiral model to produce USPs, the tidal dissipation in USP hosts needs to be efficient, but the population analysis on stellar kinematic ages seems to suggest otherwise (Hamer & Schlaufman 2020).

**2.1.4. Radius valley.** An important discovery in the field of exoplanets in recent years is the radius valley, which refers to a region in the period–radius plane at radii  $R_p \sim 2 R_{\oplus}$  and periods of  $\sim 3\text{--}30$  days (Fulton et al. 2017, Van Eylen et al. 2018, Fulton & Petigura 2018). This radius valley is visible in our statistical sample (see **Figure 2**). The position of the valley in radius is reported to decrease with the orbital period (Van Eylen et al. 2018) and increase with the stellar



**Figure 4**

(a) The zoom-in view of the *Kepler* planets in our sample centered at the radius valley. The y axis shows the rescaled radius  $\tilde{R}_p \equiv R_p(P/10 \text{ days})^{-g}(M_*/M_\odot)^{-b}$  (see Equation 12), and we adopt the best-fit  $g = -0.09$  (Van Eylen et al. 2018) and  $b = 0.26$  (Berger et al. 2020a). The black box marks the boundary within which planets are used to derive the intrinsic radius distribution. (b) The intrinsic distribution of the rescaled radius  $\tilde{R}_p$ . The radius gap, highlighted in the gray band, is most prominent when both period and stellar mass dependences are taken into account.

**mass** (Wu 2019, Berger et al. 2020a). The two dependences can be parameterized as

$$\frac{R_p}{R_{\text{valley}}} = \left( \frac{P}{10 \text{ days}} \right)^g \left( \frac{M_*}{M_\odot} \right)^b. \quad 12.$$

The valley position at orbital period  $P = 10$  days and host mass  $M_* = M_\odot$  is found to be  $R_{\text{pvalley}} = 1.9 \pm 0.2 R_\oplus$  and the slope quantifying the period dependence is  $g = -0.09^{+0.02}_{-0.04}$  (Van Eylen et al. 2018). The slope quantifying the stellar mass dependence is  $b = 0.26^{+0.21}_{-0.16}$  (Berger et al.

2020a). With the above relation one can then highlight the radius valley by rescaling the radius to  $\tilde{R}_p \equiv R_p(P/10 \text{ days})^{-g}(M_*/M_\odot)^{-b}$ . **Figure 4a** illustrates our sample in this rescaled radius (with  $g = -0.09$  and  $b = 0.26$ ) versus orbital period plane. We also show the intrinsic distribution of the rescaled radius in **Figure 4b** for planets with  $\tilde{R}_p$  in the range  $1\text{--}4 R_\oplus$  and  $P$  in the range of  $3\text{--}30$  days. Our choice of the period upper boundary is motivated by **Figure 2**: Beyond  $\sim 30$  days the number of detections in the relevant region and thus the statistical power drop significantly. The peak-to-dip contrast in our radius distribution is not as significant as that shown by Fulton et al. (2017) and Fulton & Petigura (2018). In particular, our rescaled radius distribution does not show an obvious single peak at  $\tilde{R}_p < R_{\text{pvalley}}$ . We have done computations with the same period range as used in those studies and confirm that our specific choice of the period range is not the cause of this difference. One possible reason is the different statistical methods used to infer the occurrence rate: As discussed in Section 1.2, the IDEM approach used by Fulton et al. (2017) and Fulton & Petigura (2018) tends to underestimate the frequency at low-sensitivity regions ( $R_p \sim R_\oplus$ ). The fact that the radius distribution does not seem to decrease at sub-Earth sizes suggests the presence of many undiscovered sub-Earths. The broader radius distribution may also imply that the planetary mass distribution is not as narrowly peaked as some previous studies inferred (e.g., Wu 2019).

The leading theory for the radius valley is the atmospheric evaporation driven by high-energy photons from the host star (photoevaporation) (Lopez & Fortney 2013; Owen & Wu 2013, 2017; Jin et al. 2014). In fact, the existence of the radius valley at approximately the discovered position had been predicted years before its discovery (Lopez & Fortney 2013, Owen & Wu 2013; see a historic overview in Owen 2019), which is exceptional in exoplanetary science. The photoevaporation of the atmosphere is thought to mostly take place during the early ages of the system when the star emits a higher fraction of its total luminosity at high energy ( $\lesssim 100$  Myr; e.g., Jackson et al. 2012, Tu et al. 2015, but also see King & Wheatley 2021). For close-in ( $\sim 3\text{--}30$  days) planets with core masses of a few Earth masses, the high-energy radiation is sufficient to unbind the entire hydrogen/helium atmosphere if its initial mass fraction is below some critical value (a few percent; Owen & Wu 2017). The radius valley thus emerges, separating planets with and without extended atmospheres (Lopez & Fortney 2013; Owen & Wu 2013, 2017). The observed period and stellar mass dependences can also be well explained by photoevaporation. As the orbital period increases and/or the host mass decreases, the amount of high-energy radiation the planet receives decreases, and thus the valley moves to smaller radii (Owen & Wu 2017, Wu 2019).<sup>5</sup> We refer interested readers to Owen (2019) for a comprehensive review on the photoevaporation mechanism.

According to the photoevaporation theory, the properties (e.g., location and shape) of the radius valley depend on the underlying planetary properties, especially distributions of the core mass, core composition, and atmospheric mass fraction (Lopez & Fortney 2013, Owen & Wu 2013). Therefore, the observed radius valley opens up a venue to statistically infer the properties of close-in low-mass planets at birth (Owen & Wu 2017, Jin & Mordasini 2018, Wu 2019, Rogers & Owen 2021). Assuming that photoevaporation is the underlying mechanism, these studies collectively point to a typical core mass of a few Earth masses, a core composition similar to that of the Earth (i.e., rich in silicate/iron and poor in water/ice), and a typical atmosphere mass fraction at birth of a few percent. These inferred properties have important implications for the formation and migration history of these close-in planets (see Section 4).

Although photoevaporation has seen its success in predicting and explaining the radius valley, alternative theories exist that can also explain the observed valley (e.g., Ginzburg et al. 2018, Lee & Connors 2021), of which the core-powered mass-loss mechanism is considered the main competing theory. Unlike photoevaporation, the energy source for atmosphere stripping in a core-powered mass-loss mechanism is the internal luminosity of the cooling core, and this process is expected to operate on much longer timescales ( $\sim 1$  Gyr) (Ginzburg et al. 2018). The observed period and stellar mass dependences of the radius valley (Equation 12) are also consistent with this mechanism (Gupta & Schlichting 2019, 2020). Similar to photoevaporation, the core-powered mass-loss mechanism also supports that the close-in low-mass planets have predominantly rocky cores with low water-ice fractions (Gupta & Schlichting 2019).

Attempts have been made to identify which of the two mechanisms discussed above is more responsible for the observed features. These studies made use of either the different stellar mass or age dependences of the two mechanisms (e.g., Hirano et al. 2018, Berger et al. 2020a). However, the currently available data provide no conclusive result to distinguish between the two. Larger samples and/or more precise measurements of stellar properties will be needed.

---

<sup>5</sup>Although later-type stars have higher fractions of the total luminosity emitted in higher energy ( $\propto M_{\star}^{-3}$ ; Lopez & Rice 2018) and remain active for a longer period of time, these lower-mass stars have much lower total luminosities ( $\propto M_{\star}^4$  for solar and later-type stars). The lifetime-integrated high-energy radiation at a certain orbital separation is shown to decrease with decreasing stellar mass (see McDonald et al. 2019, their figure 4).



## 2.2. Mutual Inclinations and the Intrinsic Multiplicity

The mutual inclination distribution of planets in multiplanet systems conveys important information on the formation and dynamical evolution of planetary systems. However, currently employed detection techniques are usually incapable of directly measuring mutual inclinations. This is particularly true for RV and microlensing. The transit technique is strongly biased toward (nearly) coplanar systems. Nevertheless, advancements have made it possible to statistically infer the mutual inclination distribution from the *Kepler* data.

The key issue in constraining the mutual inclination distribution with transit is the strong degeneracy with the intrinsic multiplicity (e.g., Lissauer et al. 2011, Tremaine & Dong 2012). Specifically, with the observed multiplicity function of transit alone one cannot distinguish between high-multiplicity systems with large mutual inclinations and low-multiplicity systems with small mutual inclinations. We therefore combine mutual inclinations and intrinsic multiplicity in the same discussion.

Before discussing the statistically inferred mutual inclinations, we briefly overview a handful of systems with measured large mutual inclinations. By combining HST astrometry and ground-based RV measurements, McArthur et al. (2010) measured the mutual inclination between two of the three planetary companions in the Upsilon Andromeda system to be about 30 deg. Mills & Fabrycky (2017) performed photodynamical modeling of the transit timing variation (TTV) and transit duration variation (TDV) signals of the Kepler-108 system and found the mutual inclination to be  $\Delta I = 24_{-8}^{+11}$  deg between the two transiting planets. The pi Mensae system, which hosts a long-period giant planet and a TESS transiting super Earth (Gandolfi et al. 2018, Huang et al. 2018), is reported to have significant mutual inclinations ( $\sim 30$ – $150$  deg) from joint analyses of the *Hipparcos* and *Gaia* DR2 astrometry (Damasso et al. 2020, De Rosa et al. 2020, Xuan & Wyatt 2020). Additionally, some USP systems have also been determined to have large mutual inclinations (e.g., Dai et al. 2018). More planetary systems with large mutual inclinations are expected to be found in the upcoming years, especially with *Gaia*'s capability to determine the 3D orbital configurations (Perryman et al. 2014).

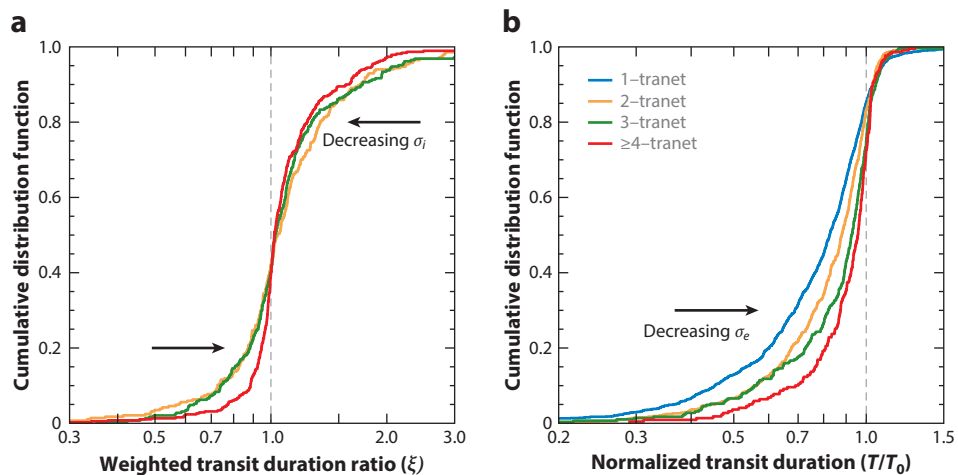
**2.2.1. The weighted transit duration method.** A popular method to statistically infer the mutual inclination of *Kepler* multiplanet systems makes use of the ratio of transit chord lengths (Steffen et al. 2010),

$$\xi \equiv \frac{T_{\text{in}} P_{\text{in}}^{-1/3}}{T_{\text{out}} P_{\text{out}}^{-1/3}} = \sqrt{\frac{(1 + r_{\text{in}})^2 - b_{\text{in}}^2}{(1 + r_{\text{out}})^2 - b_{\text{out}}^2}}. \quad 13.$$

The subscripts in and out denote values of the inner and the outer transiting planets, respectively. Here,  $T$  measures the time from the first to the last contact points of transit,  $r$  is the planet-to-star radius ratio, and  $b$  is the transit impact parameter. As both  $T$  and period  $P$  are precisely measured from transit data (Seager & Mallén-Ornelas 2003), the parameter  $\xi$  is well determined from observations. The last expression in Equation 13 is used to construct the  $\xi$  distribution from models with assumed mutual inclination distributions. When two transiting planets are exactly coplanar, the ratio  $b_{\text{in}}/b_{\text{out}} = a_{\text{in}}/a_{\text{out}} = (P_{\text{in}}/P_{\text{out}})^{2/3}$  is precisely measured, and thus the parameter  $\xi$  only concerns one poorly constrained fiducial parameter (either  $b_{\text{in}}$  or  $b_{\text{out}}$ , since both  $r_{\text{in}}$  and  $r_{\text{out}}$  are reasonably well measured). The distribution of  $\xi$  for coplanar systems is thus expected to narrowly peak at unity. In practice, the observed distribution is not so narrow, because of the introduction of the mutual inclination (see **Figure 5a**).<sup>6</sup> Applying this weighted transit duration

<sup>6</sup>The orbital eccentricity  $e$  in principle also affects the  $\xi$  distribution, but its contribution is relatively minor and thus  $e$  cannot be well constrained with this method (Fabrycky et al. 2014).





**Figure 5**

(a) The cumulative distribution functions (CDFs) of the weighted transit duration ratio,  $\xi$  (Equation 13), for different transit multiplicities. Larger transit multiplicities tend to have narrower  $\xi$  distributions, suggesting smaller mutual inclination dispersions  $\sigma_i$ . (b) The CDFs of the normalized transit duration,  $T/T_0$  (Equation 15), for different transit multiplicities. Here,  $T_0$  is the transit duration for a circular and coplanar orbit. Larger transit multiplicities have narrower  $T/T_0$  distributions, indicating smaller eccentricity dispersions  $\sigma_e$ .

method to large samples of *Kepler* planet pairs, Fang & Margot (2012) and Fabrycky et al. (2014) found that the mutual inclinations between transiting planets in the *Kepler* multiplanet systems could be well described by a Rayleigh distribution with dispersion of a few degrees ( $\lesssim 3\text{--}5$  deg). This has been frequently interpreted as multiplanet systems being nearly coplanar. However, with the use of only transiting planet pairs, which preferentially have small mutual inclinations, the weighted transit duration method cannot well determine the higher end of the mutual inclination distribution. As an extreme case, even the isotropic distribution of orbital inclinations cannot be reliably ruled out with the use of transit data alone (Tremaine & Dong 2012).

**2.2.2. Kepler dichotomy.** To recover the true mutual inclination distribution, one needs to break its strong degeneracy with intrinsic multiplicity. The first attempt was carried out by Lissauer et al. (2011). The authors tried different functional forms for the intrinsic multiplicity distribution (uniform, Poisson, and exponential) as well as for the mutual inclination distribution (uniform and Rayleigh; see also Sandford et al. 2019). By modeling the intrinsic multiplicity as a uniform (or Poisson) distribution and the mutual inclination as a Rayleigh distribution, Lissauer et al. (2011) were able to find matches to all observed transit multiplicities except the transit singles. Specifically, their models would underpredict the number of systems with only one transiting planet by nearly 50%. This signals the failure of their simplified model. Nevertheless, this feature was picked up by many others and phrased as the evidence for two distinct populations of planetary systems (the so-called “*Kepler* dichotomy”): In one population, planetary systems have small mutual inclinations and relatively compact configurations, whereas in the other population planetary systems have either only one planet or at least two largely mutually inclined planets (e.g., Johansen et al. 2012, Ballard & Johnson 2016, Mulders et al. 2018, He et al. 2019). Taking the *Kepler* sample as a whole, in terms of distributions of many properties of stars (e.g., stellar mass, metallicity) and planets (e.g., period), transit singles and transit multis are statistically consistent with being drawn

from the same parent population (e.g., Xie et al. 2016, Munoz Romero & Kempton 2018, Weiss et al. 2018a, Zhu et al. 2018b), suggesting that they probably have the same origin.

Although modeling the mutual inclination as a Rayleigh distribution (or more generally, Fisher distribution; Tremaine & Dong 2012, Zhu et al. 2018b) seems a reasonable choice (see also Tremaine 2015), the proper functional form for the intrinsic multiplicity distribution remains an open question. Nevertheless, it is certainly oversimplified to assume that all planetary systems have the same number of planets (e.g., Ballard & Johnson 2016, Mulders et al. 2018). Having a Poisson distribution for the intrinsic multiplicity (e.g., Lissauer et al. 2011, He et al. 2019, Sandford et al. 2019, Zink et al. 2019) is likely not justified either. The underlying assumption behind the Poisson distribution is that occurrences and properties of individual planets around the same host are independent from each other. Although it has not been proved invalid for *Kepler* planets, there is emerging evidence that the presence and properties of planets inside the same system may be correlated because of the shared formation environment and/or host properties (see Sections 1.1, 2.4, and 2.5). Furthermore, the exponential or power-law (i.e., Zipfian distribution; Sandford et al. 2019) forms can be securely ruled out. These distributions predict overly abundant intrinsic single-planet systems, which is not supported by TTV observations (e.g., Ford et al. 2011).

Given the strong degeneracies, disentangling the intrinsic multiplicity function and the mutual inclination distribution therefore requires external information. To this end, Tremaine & Dong (2012) developed a general statistical framework to account for observational biases of different techniques. Applying their method to planetary systems from *Kepler* and RV, Tremaine & Dong (2012) found that the mean mutual inclination dispersion, which was assumed to be the same for all multiplicities, should be  $\lesssim 5$  deg and that the intrinsic multiplicity function could not be constrained. See Figueira et al. (2012) for a different attempt in combining *Kepler* and RV data.

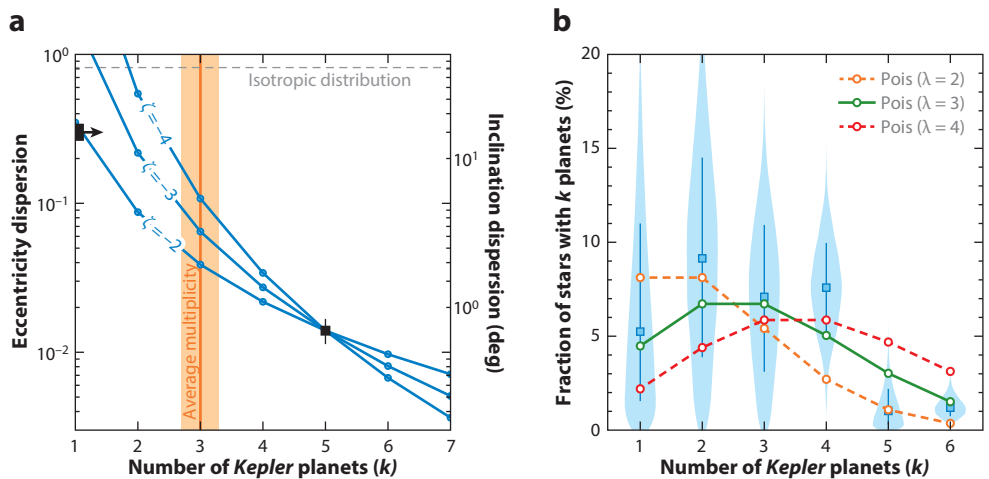
Tremaine & Dong (2012) also pointed out an observational feature that was difficult for their models to explain. As originally noticed by Ford et al. (2011), the fraction of systems showing TTV signals does not seem to vary significantly with the transit multiplicity, except perhaps for very high ( $\geq 4$ ) multiplicities (see also Xie et al. 2014). A similar feature also shows up in later large and uniform TTV searches, which consistently found that nearly half of the TTV detections were from systems with only one transiting planet (Holczer et al. 2016, Ofir et al. 2018). This indicates that planets in transit singles have almost the same probability to show TTV signals as planets in transit multis.

**2.2.3. Multiplicity-dependent mutual inclinations.** The assumption that the mutual inclination distribution is independent of the intrinsic multiplicity may not be valid. With all else being equal, the critical mutual inclination for long-term instability is probably dependent on the number of planets in the system (e.g., Pu & Wu 2015; see also Section 2.4.2). Observationally, one also finds that the distribution of the  $\xi$  parameter appears statistically different for different transit multiplicities. As shown in **Figure 5a**, lower transit multiplicities have broader  $\xi$  distributions that are suggesting larger mutual inclinations (see also He et al. 2020).

Zhu et al. (2018b) introduced the following relation between the mutual inclination dispersion,  $\sigma_i$ , and the intrinsic multiplicity (within the *Kepler* window),  $k$ :

$$\sigma_i(k) = 0.8^\circ \left( \frac{k}{5} \right)^\xi. \quad 14.$$

They applied the statistical framework of Tremaine & Dong (2012) and combined the transit and TTV statistics to infer the intrinsic multiplicity and mutual inclination distributions. TTV, as a detection technique (Agol et al. 2005, Holman & Murray 2005), applies to the same population



**Figure 6**

(a) Distributions of eccentricity and mutual inclination dispersions as functions of the intrinsic multiplicity. Here, the *Kepler* window is roughly the region above the gray solid curve in **Figure 2**. The relation  $\sigma_e = \sigma_i$  is assumed (see Section 2.3.2). The  $\zeta$  parameter quantifies the strength of the correlation with the intrinsic multiplicity (Equation 14). The orange band denotes the inferred average multiplicity  $\bar{m}_p$  for *Kepler* systems. The eccentricity dispersion inferred from transit singles (i.e.,  $k \geq 1$ ) and the mutual inclination dispersion constraint from systems with five *Kepler* transiting planets are shown as black squares. (b) The inferred intrinsic multiplicity vector and the associated probability distribution functions. The fraction of Sun-like stars with more than seven planets in the *Kepler* window is limited to  $<2.2\%$  (95% upper limit). The medians and the 16–84% ranges of individual components are denoted with squares and error bars, respectively. Poisson distributions with different values of the mean parameter  $\lambda$  are shown for reference. Both plots are adapted from Zhu et al. (2018b) with permission of the AAS.

of planetary systems as transit, and thus the combination of TTV and transit is free from many assumptions and selection biases (compared to the use of RV; e.g., Tremaine & Dong 2012). Zhu et al. (2018b) found that the intrinsic multiplicity and the mutual inclination dispersion should be strongly correlated, with  $-4 < \zeta < -2$  at the  $2\sigma$  confidence level (see **Figure 6a** for an illustration). In other words, systems with fewer planets are dynamically hotter. This result also points to large mutual inclinations ( $\gtrsim 10$  deg) for 2-planet and 3-planet systems. A recent work by He et al. (2020) found a qualitatively similar (although statistically different) result with a best-fit  $\zeta = -1.7$  from modeling a collection of *Kepler* statistics (including transit multiplicities, the period distribution, period ratio distribution, etc.) and imposing the angular momentum deficit stability criterion (Laskar 1997, Laskar & Petit 2017) in simulated planetary systems. It is also worth noting that such a relation is steeper than the similar relation inferred from RV eccentricities (Limbach & Turner 2015), ergodic models (Tremaine 2015), or the extrapolations of the empirical stability boundary (e.g., Pu & Wu 2015).

Zhu et al. (2018b) also reported constraints on the intrinsic multiplicity vector, which is reproduced in **Figure 6b**. Although the individual components of the multiplicity vector are not well constrained, the summed fraction is well measured to be  $30 \pm 3\%$  and does not rely on many assumptions like the other measurements do (see Zhu et al. 2018b, their section 5.1; see also Section 1.1). The resulting average multiplicity in the *Kepler* parameter space is  $\bar{m}_p = 3.0 \pm 0.3$ . This serves as a lower bound on the average multiplicity in the inner ( $\lesssim 1$  AU) region, as smaller planets below the detection threshold of *Kepler* are unconstrained.

## 2.3. Eccentricity Distribution

Similar to mutual inclinations, orbital eccentricities also provide important information on the formation and dynamical evolution of planetary systems. Here, we focus on the eccentricity results from the *Kepler* sample. Readers can find discussions about eccentricities from RV by Winn & Fabrycky (2015, their section 3.1).

The majority of the eccentricity measurements of individual *Kepler* planets were made through modeling the TTV and TDV signals (e.g., Lithwick et al. 2012; Wu & Lithwick 2013; Hadden & Lithwick 2014, 2017). These studies have found that the eccentricities of *Kepler* planets in near-resonance pairs are typically small, with a Rayleigh dispersion of up to a few percent. However, the planets selected for such dynamical modelings are probably a biased sample, and thus the derived eccentricity distribution may not be representative of the more general population.

**2.3.1. The transit duration method.** The transit duration (between the first and the fourth contact points) is given by<sup>7</sup>

$$\frac{T}{T_0} = \sqrt{(1+r)^2 - b^2} \frac{\sqrt{1-e^2}}{1+e \sin \omega}. \quad 15.$$

Parameters  $r$ ,  $b$ , and  $T$  are the same as those in Equation 13, and  $\omega$  is the argument of periapsis. The quantity  $T_0$  measures the transit duration between the first (second) and the third (fourth) contact points of a planet with the same period but circular ( $e = 0$ ) and edge-on ( $b = 0$ ) orbit and is related to the mean density of the host star,  $\rho_*$ , via

$$T_0 \equiv \frac{R_* P}{\pi a} = 13 \text{ h} \left( \frac{P}{\text{year}} \right)^{1/3} \left( \frac{\rho_*}{\rho_\odot} \right)^{-1/3}. \quad 16.$$

With known parameters from transit modeling ( $b$ ,  $r$ ,  $P$ , and  $T$ ) and the nuisance parameter  $\omega$  assumed to follow a uniform distribution, the quantity  $T/T_0$  can be used to constrain the statistical distribution of  $e$ , provided that the stellar mean density is precisely measured (Ford et al. 2008). With other parameters being the same, larger eccentricities lead to broader distributions of the  $T/T_0$  ratio (see **Figure 5b**). The successful application of this method heavily depends on the accurate characterizations of the host stars. As a result, early attempts to study the *Kepler* sample were all limited by the systematic uncertainties in the stellar properties (e.g., Moorhead et al. 2011, Kane et al. 2012, Plavchan et al. 2014).

**2.3.2. Multiplicity-dependent eccentricity distribution.** Van Eylen & Albrecht (2015) applied a variant of the transit duration method to a carefully selected sample of *Kepler* multiplanet systems whose host stars were precisely characterized via asteroseismology. These authors found that the eccentricities of planets in their sample could be well described by a Rayleigh distribution with  $\sigma_e \approx 0.05$ . Using accurate spectroscopic stellar parameters from LAMOST, Xie et al. (2016) found similar nearly circular orbits for planets in the *Kepler* multis, and they reported a much larger eccentricity dispersion ( $\sigma_e \approx 0.3$ ) for *Kepler* planets in systems with single transiting planets. Both results have been confirmed by later works (Mills et al. 2019, Van Eylen et al. 2019).

The multiplicity-dependent eccentricity distribution goes beyond the single versus multiple bifurcation. This is demonstrated in **Figure 5b**, where we show the cumulative distributions of the  $T/T_0$  ratios derived from our planet sample for different transit multiplicities. Here, we have

<sup>7</sup>Note that our definition of the transit duration follows that of Seager & Mallén-Ornelas (2003) and is different from that of Winn & Fabrycky (2015). The latter measures the duration between two points, where the planetary center sits on the edge of the projected stellar surface (see Winn 2010, his figure 2).

used the stellar mean densities from isochrone fits by Berger et al. (2020b) and the values of  $T$  from the *Kepler* DR25 Markov chain Monte Carlo chains (Hoffman & Rowe 2017). As **Figure 5b** shows, the distribution of the  $T/T_0$  ratio becomes narrower with increasing transit multiplicities. As the transit multiplicity can be viewed as a rough proxy for the intrinsic planet multiplicity, it suggests that planetary systems with more planets have smaller eccentricity dispersions. This is also qualitatively consistent with studies of the RV planets (Limbach & Turner 2015, Zinzi & Turrini 2017). Based on observations of the Solar System and the general expectation that the dispersions of orbital eccentricity and mutual inclination are proportional to each other (Ida et al. 1993, Tremaine & Dong 2012, Xie et al. 2016), we may use the same relation between intrinsic multiplicity and mutual inclination dispersion (Equation 14) for the relation between intrinsic multiplicity and orbital eccentricity dispersion. The multiplicity-dependent eccentricity dispersion is also shown in **Figure 6a** (see also He et al. 2020).

The large eccentricities and mutual inclinations of *Kepler* low-multiples have important theoretical implications. The largest eccentricity that can be achieved via scatterings among small *Kepler* planets themselves can be roughly estimated as

$$e_{\max} \sim \frac{v_{\text{esc}}}{v_{\text{orb}}} \approx \sqrt{\frac{2m_p a}{M_\star R_p}} = 0.15 \left( \frac{m_p/M_\star}{10^{-5}} \right)^{1/2} \left( \frac{a}{0.1 \text{ AU}} \right)^{1/2} \left( \frac{R_p}{2 R_\oplus} \right)^{-1/2}. \quad 17.$$

Here,  $v_{\text{esc}}$  and  $v_{\text{orb}}$  are the surface escape velocity and orbital velocity of the planet, respectively. The evaluation takes the typical values of a *Kepler* planet. Although the above scaling relation bears some significant uncertainties, the large eccentricities ( $\sigma_e \approx 0.3$ ) and mutual inclinations ( $\sigma_i \gtrsim 10^\circ$ ) observed in the low-multiplicity planetary systems are probably on the high end of the distribution. This suggests that these planetary systems may have undergone significant dynamical interactions among the inner planets themselves. Alternatively, other mechanisms may have been invoked to excite eccentricities and mutual inclinations to values larger than what the self-scatterings can achieve. One promising mechanism is the interaction between the inner system and the outer massive planets (e.g., Johansen et al. 2012, Huang et al. 2017, Pu & Lai 2020, and references therein). We return to this point in Section 3.2.

## 2.4. Intrasystem Variation

The intrasystem variation, which is about the relative properties of planets around the same host, is useful in constraining the formation and evolution processes of planetary systems. It also concerns the statistical inference of exoplanets in general: In some statistical studies, planet detections from the same star are treated as independent events (see Sections 1.2 and 2.2); in some others, specific assumptions about the relative properties of planets in multiplanet systems must be made when synthetic systems are generated (e.g., Mulders et al. 2018, He et al. 2019). The derived statistics to some extent are subject to the validity of such assumptions.

**2.4.1. Peas in a pod?** Transiting planets in the same *Kepler* multiplanet systems preferentially have similar sizes. This feature has been noticed since the early days of the *Kepler* mission (Lissauer et al. 2011, Ciardi et al. 2013). Follow-up observations that provided improved characterizations of the host stars enabled further studies that tried to understand the nature of this feature (Weiss et al. 2018b, He et al. 2019, Murchikova & Tremaine 2020, Weiss & Petigura 2020, Zhu 2020). In particular, Weiss et al. (2018b) quantified the correlation between sizes of neighboring planets around the same host in their sample. To check the statistical significance of this correlation, they generated synthetic systems by randomly drawing planetary radii from the observed size distribution and then performed the same correlation test. The size correlations in their synthetic systems

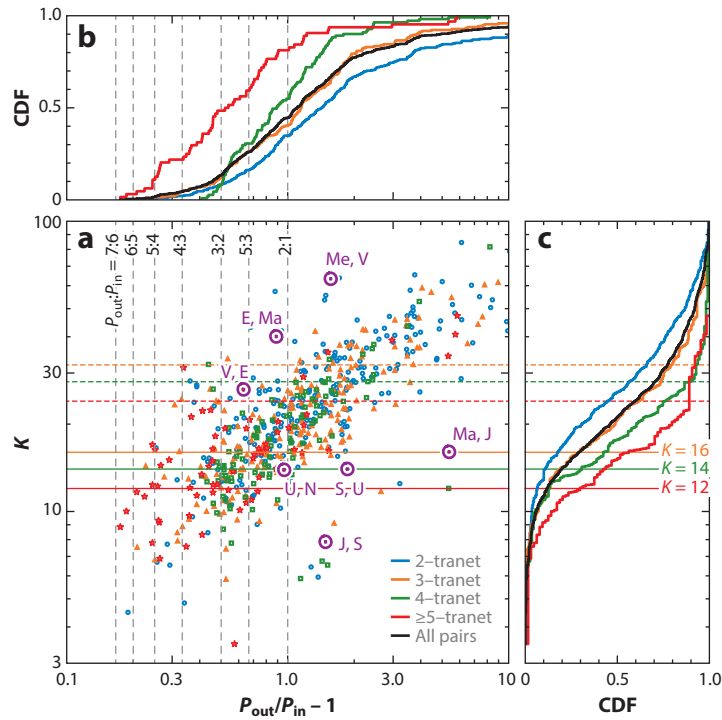
were much weaker than what they saw in real systems, and thus they concluded the pattern was astrophysical. Based on this and on a similar result on the spacings between planets, Weiss et al. (2018b) concluded that planets in *Kepler* multiplanet systems have similar sizes and regular spacings, a pattern they termed “peas in a pod” (see also Millholland et al. 2017 for a similar claim about *Kepler* planet masses). A later study by He et al. (2019) reached a similar conclusion. According to these authors, planetary systems that contain clusters of planets whose sizes and orbital periods are correlated produce a better match to the observed *Kepler* systems in terms of the joint statistics of transit depth distribution, period distribution, period ratio distribution, etc.<sup>8</sup> Different opinions exist about the nature of the observed correlations. Zhu (2020) pointed out a detection bias that was underestimated in the statistical method of Weiss et al. (2018b). Because small planets can be detected around bright and quiet stars, whereas large planets are only detectable around faint or noisy stars, the same transit detection threshold (i.e., a fixed S/N) naturally leads to varying planetary size thresholds in different systems. This, combined with the fact that smaller planets are more abundant, naturally leads to a size correlation in the observed transit pairs (see also Murchikova & Tremaine 2020). However, it appears that the apparent correlation in planetary sizes is too strong to be explained entirely by this detection bias alone (Zhu 2020).

Another factor that has not been fully explored is the contribution of the planets that are missing, due to large impact parameters or subthreshold values of transit S/N, in known *Kepler* multiplanet systems. Our Solar System is an excellent example to demonstrate this point. The four outer giant planets would not likely be detected by a transit mission similar to *Kepler* because of their long orbital periods. Of the four terrestrial planets, Mercury and Mars are almost impossible to detect in transit due to their small sizes. Therefore, a *Kepler*-like mission would, if possible at all, most likely detect the Venus–Earth planet pair, which shows very similar sizes ( $0.95 R_{\oplus}$  versus  $1 R_{\oplus}$ ) and masses ( $0.82 M_{\oplus}$  versus  $1 M_{\oplus}$ ). However, this level of similarity is not representative among the Solar System planet pairs.

The physical interpretation of the size correlation (if any) is also unclear. One interpretation is that planets “know” about their siblings, namely the formations of two neighboring planets are directly correlated (e.g., Kipping 2018, Mulders et al. 2018, He et al. 2019, Sandford et al. 2019, Gilbert & Fabrycky 2020). Another interpretation is that planets “know” about the system and the environment in which they formed, namely the formations of planets in the same system are all related to some global properties (Murchikova & Tremaine 2020). In this latter case, the apparent correlation between planetary sizes is only a projection of the correlation between the individual planets and the host star (or the birth disk). This latter interpretation has some observational evidence. For example, the planet distribution is shown to depend on the orbital period (see Section 2.1) and stellar properties (see Section 2.5). Murchikova & Tremaine (2020) demonstrated with a toy model that the observed size correlation can be well reproduced if the planets “know” about the host star but do not “know” about their neighbor planets.

**2.4.2. Orbital spacings.** The relative positions of planets in *Kepler* multiplanet systems have also drawn lots of interest. The majority of the early studies focused on the period ratio distribution. As shown in **Figure 7**, *Kepler* systems contain very few planet pairs near/in low-order mean-motion resonances (see also Lissauer et al. 2011, Fabrycky et al. 2014). This is in contrast with earlier RV results showing that a substantial fraction of well-characterized multiplanet

<sup>8</sup>The clustered model of He et al. (2019) has more free parameters than their nonclustered model. However, the authors did not perform model comparisons to justify the introduction of more flexibilities. See Zhu (2020) for more discussion.



**Figure 7**

Spacings between the apparently adjacent planets in *Kepler* multiplanet systems, with different colors indicating planet pairs from different transit multiplicities. The  $x$  axis of panel *a* shows the spacing in terms of the orbital period ratio, and panel *b* shows the corresponding CDFs. A few example period commensurabilities are indicated in both panels. The  $y$  axis of panel *a* shows the spacing in terms of the mutual Hill radii (Equation 18), and panel *c* shows the corresponding cumulative distributions. The stability thresholds for 3-tranet, 4-tranet, and 5-tranet systems, derived according to Equation 20, are indicated with solid horizontal lines. Values corresponding to twice the thresholds are also shown as dashed horizontal lines. The code *Forecaster* from Chen & Kipping (2017) is used to predict the planet mass based on the planetary radius, and we have revised the upper mass limit to  $10^3 M_{\oplus}$  ( $\sim 3 M_J$ ) to avoid masses beyond the planetary regime. Solar System planet pairs are also indicated for reference. Abbreviations: CDF, cumulative distribution function; E, Earth; J, Jupiter; Ma, Mars; Me, Mercury; N, Neptune; S, Saturn; tranet, transiting planet; U, Uranus; V, Venus.

systems contain pairs of giant planets close to mean-motion resonances (e.g., Wright et al. 2011). We refer to Section 4.1 for the theoretical implications of this feature. Additionally, the asymmetry around exact period commensurabilities has also attracted lots of attention (Fabrycky et al. 2014), and we refer interested readers to the fairly comprehensive overview by Terquem & Papaloizou (2019) for this particular issue (see also the recent development described by Millholland & Laughlin 2019). This review focuses on the dynamical compactness of the *Kepler* multiplanet systems, which concerns the long-term stability and thus the dynamical evolution.

When the stability of the planetary system is concerned, the orbital spacing between planets is usually expressed in the dimensionless parameter  $K$ :

$$K \equiv \frac{a_{\text{out}} - a_{\text{in}}}{R_H}; \quad R_H \equiv \frac{a_{\text{in}} + a_{\text{out}}}{2} \left( \frac{m_{\text{in}} + m_{\text{out}}}{3M_{\star}} \right)^{1/3}. \quad 18.$$



Here,  $R_H$  is called the mutual Hill radius,  $M_*$  is the mass of the host star, and  $a_{\text{in}}$  ( $a_{\text{out}}$ ) and  $m_{\text{in}}$  ( $m_{\text{out}}$ ) are the semimajor axis and the mass of the inner (outer) planet, respectively. For two-planet systems, the condition for the long-term stability (and thus instability) has been well understood theoretically, and the instability arises when there are mean-motion resonance overlaps (Wisdom 1980, Deck et al. 2013, Hadden & Lithwick 2018). For systems with more than two planets, we lack a good theoretical understanding of the origin of the dynamical instability (see attempts by Chambers et al. 1996, Zhou et al. 2007, Quillen 2011, Yalinewich & Petrovich 2020). Nevertheless, numerical studies have shown that the timescale before which close encounter occurs between planets,  $t$ , scales exponentially with the initial spacing  $K$  (Chambers et al. 1996). Details of this scaling relation depend on factors such as the number of planets, planet masses, orbital eccentricities and inclinations, as well as the inhomogeneity among planets (e.g., Chambers et al. 1996, Zhou et al. 2007, Funk et al. 2010; see Pu & Wu 2015 for a recent summary).

In the context of *Kepler* planetary systems, Pu & Wu (2015) found through numerical simulations that the median spacing for stability could be approximated as

$$\langle K \rangle = 2.87 + 0.7 \log_{10} \tau + 2.4 \left[ \left( \frac{\sigma_e}{e_H} \right) + \left( \frac{\sigma_i}{4e_H} \right) \right], \quad 19.$$

where  $\tau$  is the physical timescale  $t$  scaled by the orbital period of the innermost planet,  $e_H$  is the mutual Hill radius scaled by the semimajor axis of the innermost planet, and  $\sigma_e$  and  $\sigma_i$  are the dispersions of orbital eccentricities and mutual inclinations among the planets, respectively. With the multiplicity-dependent  $\sigma_e$  and  $\sigma_i$  (Equation 14) and the typical values for *Kepler* systems ( $t \gtrsim 1$  Gyr and the innermost planet of planet-to-star mass ratio  $q \approx 10^{-5}$  at 0.1 AU), Equation 19 yields

$$\langle K \rangle \approx 10.2 + 2.2 \left( \frac{k}{5} \right)^\zeta. \quad 20.$$

With  $\zeta = -2$  (Zhu et al. 2018b, He et al. 2020), planetary systems with  $\{3, 4, 5\}$  *Kepler* planets should have critical spacings  $\langle K \rangle = \{16, 14, 12\}$ , respectively.

We apply the above stability thresholds to the multiplanet systems from Section 2.1 and discuss the limitations. After the use of Kepler’s third law, the only unknown to determine the spacing parameter  $K$  is the planet-to-star mass ratio. We estimate the planetary masses from the measured radii with the *Forecaster* code from Chen & Kipping (2017) and adopt the *Gaia* stellar masses from Berger et al. (2020b). Systems without reported stellar mass measurements are excluded. **Figure 7** illustrates the spacings between neighboring *Kepler* planets of all systems, with the systems being divided into different transit multiplicities. For transit multiplicities of 3, 4, and 5+, the majority ( $\sim 70\%$ ) of planet pairs have spacings above the corresponding stability thresholds, confirming that they are indeed (most likely) long-term stable. The remaining  $\sim 30\%$  of planet pairs, considered long-term unstable by the above empirical thresholds, are probably stable as well. Although part of this misclassification is due to the choice of fixed *Kepler* system parameters and the empirical (but sometimes unphysical) mass–radius relation (Chen & Kipping 2017), it nevertheless is a sign of the failure of the empirically determined stability criteria. In particular, these stability criteria do not consider the impact of mean-motion resonances, which can be either protective or destructive to the involved planets.

Nevertheless, by applying the empirical stability thresholds to the data one finds that the majority of *Kepler* planet pairs are not far from the empirical stability limits: The median spacing of all planet pairs is  $K \approx 20$ , and about 80–90% of planet pairs from systems with at least three transiting planets have spacings within twice the empirical stability thresholds (see **Figure 7**). These results are consistent with previous findings (e.g., Fang & Margot 2013, Pu & Wu 2015, Weiss et al.

2018b) and also suggest that for the majority of *Kepler* planet pairs there is no room for inserting another (undetected) planet in between (Fang & Margot 2013). In other words, the observed *Kepler* planets are dynamically packed. However, this does not necessarily mean that *Kepler* systems do not contain additional planets. The spaces inside the innermost and particularly beyond the outermost *Kepler* planet allow the existence of additional planets without risking instability. For example, seven planets with  $q = 10^{-5}$  are allowed per factor of 10 in semimajor axis if mutually separated by  $K = 20$ . The observed dynamically packed structure is also probably due to the selection bias that it is more difficult to detect both planets transiting to the host star for a planet pair with wider spacing.

As part of the peas-in-a-pod claim (see Section 2.4.1), the spacings between *Kepler* planets in the same multiplanet system are found to be statistically similar (Weiss et al. 2018b). However, the observed correlation in spacings is driven by a small fraction ( $\lesssim 5\%$ ) of systems containing the highest multiplicities, and the majority of systems do not show such a regular spacing pattern (Jiang et al. 2020, Zhu 2020).

## 2.5. Dependence of Planet Statistics on Stellar Properties

The distributions of the planetary systems may be correlated with the properties of their host stars. In this section, we discuss the observational constraints on such correlations.

**2.5.1. Impact of stellar companions.** Stellar companions to the planet hosts affect the *Kepler* planet statistics in several ways. In transit surveys like *Kepler*, many of them appear unresolved and dilute the transit signals, potentially leading to misclassifications and erroneous planetary parameters (Ciardi et al. 2015, Bouma et al. 2018).<sup>9</sup> Thankfully, follow-up high-resolution imaging observations have been performed for nearly all *Kepler* planet candidates (e.g., Furlan et al. 2017, Ziegler et al. 2018, and references therein). For bright targets that contain Jupiter-like transits, Santerne et al. (2016) also performed systematic RV follow-up observations and identified a significant false positive rate (55%) for Jovian planet candidates. These efforts have led to a much better understanding of the impact of transit dilution on the *Kepler* planet statistics. In particular, Furlan et al. (2017) reported that about 10% (30%) of the candidate host stars have observed companions within 1 arcsec (4 arcsec), the majority of which are fairly faint compared to the target stars. In the most likely scenario that the transit signals come from the primary stars (see, e.g., Bouma et al. 2018), the dilution effect overall only affects the planetary radii up to a few percent on average (Furlan et al. 2017). This is within the uncertainty of *Gaia*-derived radii, and thus one does not expect it to have a significant impact on the general planet statistics. However, Earth-sized planets with radii  $R_p \lesssim 2 R_\oplus$  are much more susceptible to the dilution effect, and thus the relevant statistics may suffer a more dramatic impact (Furlan et al. 2017, Bouma et al. 2018).

Besides the transit dilution effect, stellar companions can also affect the presence of planets through many dynamical processes (e.g., Artymowicz & Lubow 1994, Holman & Wiegert 1999). Very close ( $a \lesssim 0.5$  AU) stellar binaries can host circumbinary (i.e., planetary-type or P-type) planets, and over a dozen such systems have been found (see Winn & Fabrycky 2015, their section 6.2). We limit our discussions to the observational aspects of circumstellar (i.e., satellite-type or S-type) planets and the implications on planet statistics.

---

<sup>9</sup>Here, an ambient star that is not physically associated with the target is also considered a companion to the target star.

Studies based on RV and high-resolution imaging observations suggest that the existence of a close stellar-mass companion is usually associated with a lower frequency of circumstellar planets (e.g., Wang et al. 2014, 2015; Kraus et al. 2016; Ngo et al. 2016; Moe & Kratter 2019). This effect is quantified by a suppression factor  $S_{\text{bin}}$ , which is the ratio between the fraction of planet hosts with stellar companions and the fraction of field stars with the same type of companions (Kraus et al. 2016, Moe & Kratter 2019). Circumstellar planets are almost completely suppressed ( $S_{\text{bin}} \lesssim 15\%$ ) when the stellar companions are close (with separation  $a_{\text{bin}} \lesssim 10$  AU), regardless of the planetary size or observed multiplicities. Planets are nearly unaffected ( $S_{\text{bin}} \gtrsim 85\%$ ) if the stellar companions are distant ( $a_{\text{bin}} \gtrsim 100$  AU). At intermediate separations ( $\sim 10$ – $100$  AU), the suppression effect gradually decreases with the increasing separation. See Moe & Kratter (2019, their figure 3) for a compilation of observational studies and an illustration of the suppression factor  $S_{\text{bin}}$  as a function of the binary separation.

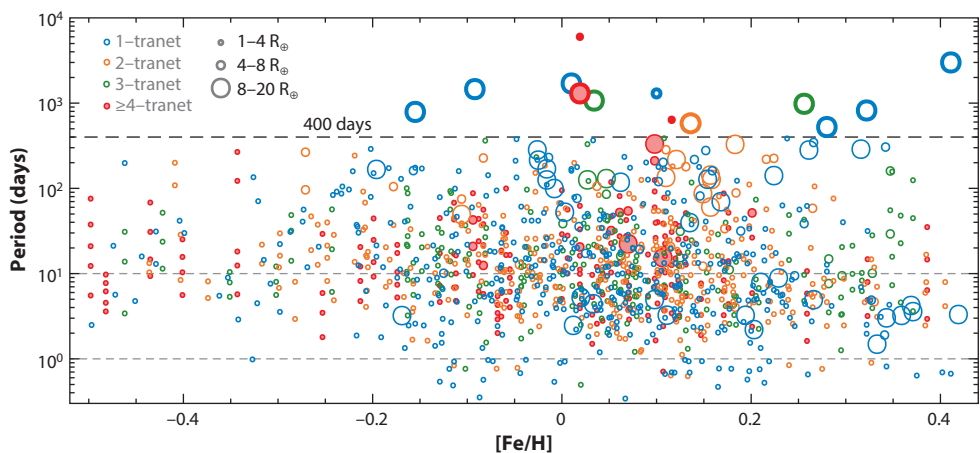
With the above suppression effect and the known binary separation distribution, one can then infer the planet-formation efficiency from the measured planetary system frequency,  $F_p$ . Moe & Kratter (2019) estimated that  $F_{\text{bin}} \approx 43\%$  of Sun-like primaries in a magnitude-limited survey like *Kepler* could not host close-in ( $\lesssim 1$  AU) planets simply because of the influence of binary companions (see also Kraus et al. 2016). If these targets are excluded from the *Kepler* statistics, one finds that the formation efficiency of close-in planets around single stars, a parameter directly related to formation theories, should be  $1/(1 - F_{\text{bin}}) = 1.8$  times higher than the fraction of stars with planets  $F_p$ . This additional factor also provides a plausible explanation to the discrepancy in hot Jupiter frequencies measured from RV and *Kepler* (see Section 2.1.1).

**2.5.2. Metallicity effect.** Under the general assumption that the bulk metallicity of the host star is correlated to the total mass of building blocks available for planet formation, it is reasonable to believe that the planet frequency and properties may be correlated with the host star metallicity. For giant planets ( $R_p \gtrsim 8 R_{\oplus}$  or  $m_p \gtrsim 0.3 M_J$ ) found by RV, it has been well established that their frequency correlates strongly with the host metallicity (e.g., Santos et al. 2001, Fischer & Valenti 2005). This giant planet–metallicity correlation lends support to the core accretion model as the leading theory for the formation of giant planets (e.g., Pollack et al. 1996, Ida & Lin 2004b). Some recent studies have also claimed that hosts of eccentric giant planets are more metal-rich than hosts of nearly circular giant planets (Dawson & Murray-Clay 2013, Buchhave et al. 2018), but stronger statistical evidence is needed to fully establish this result.

Small planets, in particular those with radii  $R_p \lesssim 4 R_{\oplus}$ , show weaker dependences on host metallicity (e.g., Sousa et al. 2008, Buchhave et al. 2012). Although many studies have focused on the dependence of the planet frequency  $\bar{n}_p$  on host metallicity (e.g., Wang & Fischer 2015, Petigura et al. 2018) and theoretical implications (e.g., Owen & Murray-Clay 2018, Lee 2019), one may argue that the planetary system frequency  $F_p$  is probably a more suitable parameter to characterize the efficiency of planet formation under such system-wide parameters like metallicity (Zhu et al. 2016, Zhu 2019). If the general planet–metallicity relation,

$$F_p \propto 10^{\gamma[\text{Fe}/\text{H}]}, \quad 21.$$

is applied, the result of Zhu (2019) suggests  $\gamma \approx 0.5$  for all *Kepler*-type planets, which is much weaker than the giant planet–metallicity correlation ( $\gamma \approx 2$ ; Fischer & Valenti 2005). The dependence is further reduced if the close binaries that show anticorrelation with stellar metallicity are excluded from the statistics (Moe et al. 2019, Kutra et al. 2020). Unlike the planetary system frequency  $F_p$ , the planet frequency  $\bar{n}_p$  does not appear to have a monotonic relation with the host metallicity. In particular, it may start declining when the metallicity is high enough (Zhu 2019). It



**Figure 8**

An illustration of the *Kepler* planetary systems in our baseline sample that have spectroscopic metallicity measurements. We use different colors to separate different observed multiplicities (out to 400 days) and different label sizes to separate planets of different sizes (small as  $1\text{--}4 R_{\oplus}$ , intermediate as  $4\text{--}8 R_{\oplus}$ , and giant as  $8\text{--}20 R_{\oplus}$ ). Cold ( $P > 400$  days) planets found by radial velocity (as tabulated in Zhu & Wu 2018) and long-period transit searches (Kawahara & Masuda 2019) are also indicated with thick circles. As the host metallicity increases, the system becomes more likely to contain giant planets at all periods and small planets at relatively close-in ( $P \lesssim 10$  days) orbits (e.g., Mulders et al. 2016, Dong et al. 2018, Petigura et al. 2018). At very high metallicities ( $[\text{Fe}/\text{H}] \gtrsim 0.2$ ), there seems to be a deficit of compact systems (with  $\geq 4$  transiting planets) and planets at intermediate orbits ( $\sim 10\text{--}400$  days). These may be related to the emerging cold giants (Zhu & Wu 2018). The median metallicity of *Kepler* field stars is  $[\text{Fe}/\text{H}] \approx -0.0$  (Dong et al. 2014b).

has been suggested that this behavior may be related to the formation of giant planets inside the same system: As the metallicity is high enough, the system has a significant probability to form giant planets, and these giants may reduce the multiplicity of the inner system because they either prohibit the formation of more small planets or dynamically remove some of the small planets out of the inner system. This scenario may also explain the increased diversity of planets around metal-rich *Kepler* hosts (Petigura et al. 2018) and the overabundant compact planetary systems around metal-poor stars (Brewer et al. 2018, Zhu & Wu 2018). **Figure 8** displays along the host metallicity  $[\text{Fe}/\text{H}]$  the *Kepler* systems with metallicity measurements in our baseline sample.

Although the stellar bulk metallicity measured in iron abundance  $[\text{Fe}/\text{H}]$  (or a mix of metals  $[\text{m}/\text{H}]$ ) is usually used in studies of the planet metallicity dependence, other elemental abundances, in particular  $\alpha$  elements and refractory elements, have also been explored for possible correlations with planet properties (e.g., Adibekyan et al. 2012, Liu et al. 2016, Teske et al. 2019). No trend has been firmly established thus far, probably due to the limited sample size, the measurement precision, and/or the impact of Galactic chemical evolution.

**2.5.3. Dependence on stellar mass.** A number of studies have also investigated the dependence of planet frequency on host mass. A theoretical possibility is that the stellar mass correlates with the total mass in the protoplanetary disk and, thus, the amount of solid materials available for planet formation. It is largely consistent with direct observations of protoplanetary disks in (sub-)millimeter wavelengths (Andrews et al. 2013, Ansdell et al. 2016), although at a fixed stellar mass the scatters of inferred disk masses remain substantial (up to an order of magnitude; Ansdell et al. 2016).

We would like to start by pointing out several potential issues. Similar to the metallicity dependence (see Section 2.5.2), the two frequencies,  $F_p$  and  $\bar{n}_p$ , can behave differently, especially for the small planets with high multiplicity rates. Second, as more massive stars also tend to be more metal-rich, one may need to carefully disentangle possible correlations between stellar mass and metallicity in the sample (e.g., Johnson et al. 2010, Kutra et al. 2020). Furthermore, the choice of the parameter used to study the correlation may matter. Although planetary radius (or mass) and orbital period (or semimajor axis) are commonly used in statistical studies, nature may prefer other physical units such as the planet-to-star mass ratio or the position of the water snow line (Hayashi 1981, Kennedy & Kenyon 2008). Last but not least, as far as the planet-formation efficiency is concerned, one must correct for the suppression effect due to close stellar binaries (see Section 2.5.1). It is established that the close binary fraction correlates with the primary mass (e.g., Duchêne & Kraus 2013), so the suppression effect is expected to affect the statistics of planets around different stellar masses differently (Moe & Kratter 2019).

The dependence of giant planets on stellar mass has been investigated in many studies with different detection methods (e.g., Johnson et al. 2007, 2010; Howard et al. 2012; Fressin et al. 2013; Nielsen et al. 2019). To avoid many of the issues listed above, here we focus on the results from long-term RV surveys, as they cover a broad range of parameter space and are nearly free of close stellar binaries. In particular, Johnson et al. (2010) analyzed a sample of 1,266 stars with at least 3-year RV observations and masses spanning from  $0.2 M_\odot$  up to their estimated  $1.9 M_\odot$  and reported a linear relation between the planet frequency<sup>10</sup> and stellar mass. This result has been widely considered as a benchmark in both theoretical and observational studies of giant planets. The higher-mass part of the sample comes from the so-called retired A-stars, and their spectroscopic mass estimates are controversial (Lloyd 2011, Schlaufman & Winn 2013, Malla et al. 2020, and references therein). Recently, the asteroseismic study by Malla et al. (2020) shows that the retired A-stars with spectroscopic masses  $> 1.6 M_\odot$  are overestimated, confirming the earlier reports by Lloyd (2011) and Schlaufman & Winn (2013). Because such stars consist of the heavier half of the Johnson et al. (2010) sample and contribute most of the statistical evidence to the reported stellar mass dependence (see Johnson et al. 2010, their figure 4), a revisit of the mass correlation will be needed. Additionally, the result of Johnson et al. (2010) is limited to the region with separation  $a < 2.5$  AU. As shown by Clanton & Gaudi (2014, 2016), after those at (slightly) larger separations are taken into account, giant planets are almost as common around M dwarfs as they are around Sun-like stars (see Section 3.3).

For small planets, the *Kepler* survey provides the best sample to study their stellar mass dependence. Studies have shown that the planet frequency,  $\bar{n}_p$ , in the *Kepler* parameter space is anticorrelated with stellar mass (e.g., Howard et al. 2012; Mulders et al. 2015a,b). Using the Berger et al. (2018, 2020b) sample with stellar effective temperature in the range of 4,000–5,000 K, we find a frequency of  $\bar{n}_p = 3.3 \pm 0.4$  for planets in the radius range of 1–20  $R_\oplus$  and period  $< 400$  days, which is a factor of  $\sim 2.7$  higher than the rate for our baseline Sun-like sample (Section 2.1). Later M-type stars have even more planets (Dressing & Charbonneau 2013, 2015). The planetary system frequency  $F_p$  is also anticorrelated with stellar mass but likely at a weaker level, due to the increased average planet multiplicity around later-type stars (e.g., Yang et al. 2020). There is some sign of increased observed multiplicity rate in our 4,000–5,000-K sample (48.2%) compared to that (42.5%) of our baseline Sun-like star sample (Section 2.1). After the correction for the suppression effect due to close stellar companions, the difference in formation efficiencies of small

<sup>10</sup>Their derived planet frequency is technically  $\bar{n}_p$ , but because of the low multiplicity rate of giant planets it closely approximates the rate  $F_p$ .

planets between single Sun-like and later-type hosts is likely further reduced, although an anti-correlation probably remains (Moe & Kratter 2019).

### 3. THE OUTER PLANET POPULATION

In the earliest stage of planet formation, the region beyond  $\sim 1$  AU is expected to contain most of the mass and the angular momentum of the protoplanetary disk. Therefore, the frequency and properties of planets in this outer region ( $\sim 1$ –10 AU) have important implications to the formation and evolution of the whole system, including the planets in the inner  $\sim 1$ -AU region. In this section, we review our current understanding of this outer planet population and discuss its connection with the inner planetary system.

We set the inner and outer boundaries at  $\sim 1$  AU partly because this is approximately the detection limit of the *Kepler* mission but also because it coincides with the position at which giant planets show a rapid rise in frequency. RV surveys have found that giant planets ( $m_p \gtrsim M_{\text{Sat}} \approx 0.3 M_J$ ) appear about five times more often at  $\sim 1$ –3 AU than they do within  $\sim 1$  AU (Cumming et al. 2008).

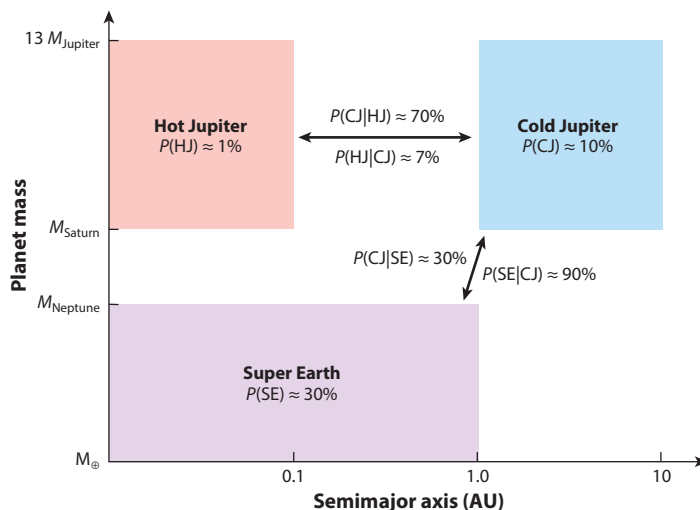
#### 3.1. Planet Frequency

RV surveys have found that cold giant planets ( $0.3$ – $13 M_J$  at  $\sim 1$ –5 AU) appear around on the order of  $\sim 10\%$  of Sun-like stars. If the giant planet distribution is modeled as a parametric function that joins single power-law distributions of mass and orbital period (Tabachnik & Tremaine 2002), the integrated rate out to  $P \approx 5.5$  years is found to be  $\bar{n}_p = 0.105$  (Cumming et al. 2008). Such a single power-law period distribution tends to overpredict the number of giant planets at wider ( $\gtrsim 10$  AU) separations. To better match the observed distribution, Fernandes et al. (2019) replaced it with a broken power law and found a potential peak at  $\sim 2$ –3 AU (see also Bryan et al. 2016). Extending their distribution function out to 100 AU, Fernandes et al. (2019) found  $\bar{n}_p = 0.27^{+0.08}_{-0.05}$  and  $0.062^{+0.015}_{-0.012}$  for planets in the mass range of  $0.1$ – $20 M_J$  and  $1$ – $20 M_J$ , respectively. If the frequency of the so-called Jupiter analogs, namely Jupiter-mass ( $\sim 0.3$ – $3 M_J$ ) planets in Jupiter-like (a few astronomical units) orbits around Sun-like hosts, are concerned, several independent studies have collectively pointed to a rate of about a few percent (e.g., Wittenmyer et al. 2016, and references therein), suggesting that planetary systems similar to our own may be relatively uncommon (see also Section 3.2). Unlike our Jupiter, a significant fraction of cold giant exoplanets are on substantially eccentric orbits with typical eccentricities  $e \sim 0.3$  (e.g., Wright et al. 2009). We refer to the review by Winn & Fabrycky (2015) for more discussion on these topics.

Although the region beyond  $\sim 1$  AU is nominally out of the reach of *Kepler*, studies have nevertheless systematically searched for and statistically studied the long-period transiting planets in the *Kepler* data (e.g., Foreman-Mackey et al. 2016, Herman et al. 2019, Kawahara & Masuda 2019). In particular, Herman et al. (2019) reported a frequency of  $\bar{n}_p = 0.7^{+0.4}_{-0.2}$  for planets with sizes of  $0.3$ – $1 R_J$  and orbital periods of  $2$ – $10$  years. The inferred radius distribution also suggests that cold Neptune-sized ( $3$ – $5 M_\oplus$ ) planets are about four times more common than cold Jupiter-sized ( $7.5$ – $11 M_\oplus$ ) ones. This is broadly consistent with the result from microlensing surveys (see Section 3.3), pointing to the potential existence of a large and unexplored low-mass planet population in the outer region.

#### 3.2. The Inner–Outer Correlation

The planetary systems inside and outside of  $\sim 1$  AU appear strongly correlated. Such a strong inner–outer correlation has important implications to the formation and evolution of the system



**Figure 9**

Correlations between inner planets and outer cold Jupiters. Although only two types of inner planets are highlighted here, they are representative of the known inner planet population: The majority of hot Jupiters do not have close neighbors (see Section 2.1.1), whereas super Earths usually reside in systems that include other types of inner planets (see Section 2.2). The unconditional probability shown here is the fraction of Sun-like stars with a specific type of planets, and the conditional probability is the fraction of Sun-like stars with a specific type of planets given that another type of planets is present in the system.

as a whole. Below, we review the observational evidence and discuss briefly the implications of this correlation. More on the latter is presented in Section 4.2. We highlight two classes of inner planets, hot Jupiters and super Earths, and discuss them separately below.

**3.2.1. Friends with close-in Jupiters.** Hot Jupiters, though usually having no detectable planetary companions in the inner region, are frequently found to have distant massive companions (Knutson et al. 2014, Bryan et al. 2016; but see also Schlaufman & Winn 2016). Both of these features are important clues to the formation and evolution of hot Jupiters, and we refer interested readers to the review by Dawson & Johnson (2018) for in-depth discussion.

For the completeness of the discussion about the inner–outer correlation, we briefly summarize here the key result of the friends of hot Jupiters search. Knutson et al. (2014) conducted a systematic RV study of the distant companions to a sample of 51 hot Jupiters and reported that each hot Jupiter should have on average  $0.51 \pm 0.10$  companions with masses of  $1\text{--}13 M_{\text{J}}$  and semimajor axes of  $1\text{--}20$  AU. This sample was reanalyzed by Bryan et al. (2016) with improved sensitivity calculations, and the companion rate was revised to  $0.70 \pm 0.08$ . Given the small fraction of systems with more than one cold companion, we take this average number to be approximately the fraction of hot Jupiter hosts with cold Jupiter companions. This fraction barely changes after we adjust to the parameter range used in this work ( $0.3\text{--}13 M_{\text{J}}$  and  $1\text{--}10$  AU) according to the planet distribution function of Bryan et al. (2016). We denote this fraction as  $P(\text{CJ}|\text{HJ})$ . Additionally, given the known fractions of Sun-like stars with hot Jupiters and cold Jupiters,  $P(\text{HJ}) \approx 1\%$  and  $P(\text{CJ}) \approx 10\%$ , respectively; the inversed conditional probability is  $P(\text{HJ}|\text{CJ}) \approx 7\%$ . This is the fraction of cold Jupiter hosts with hot Jupiters. All four fractions are shown in **Figure 9**.

Jupiter-sized planets in the inner region with known outer giant companions tend to have higher eccentricities, suggesting possible dynamical interactions in sculpting the architectures of



these systems (e.g., Bryan et al. 2016). In particular, warm ( $\sim 10\text{--}100$  days) Jupiters on significantly eccentric orbits ( $e \gtrsim 0.4$ ) have much higher chances to possess relatively close ( $\lesssim 3$  AU) Jovian companions compared to those on nearly circular ( $e \lesssim 0.1$ ) orbits (Dong et al. 2014a), and the existence of such companions is consistent with the high-eccentricity migration scenarios to form eccentric warm Jupiters (Dong et al. 2014a, Dawson & Chiang 2014, Petrovich & Tremaine 2016). By contrast, warm Jupiters on nearly circular orbits show a weaker correlation with outer giant companions, and many of these warm Jupiters are found to have nearby small planetary companions (Huang et al. 2016). These features cannot be easily reconciled in the high-eccentricity migration scenarios, suggesting that the nearly circular warm Jupiters may have been formed in situ or have undergone the disk-driven migration (e.g., Raymond et al. 2008, Hallatt & Lee 2020). We refer to Dawson & Johnson (2018) for a more comprehensive discussion on the observations and theories related to warm Jupiters.

**3.2.2. Super Earth–cold Jupiter relation.** The term super Earth has different meanings in different studies. Here, we call a planet super Earth if its mass (or radius) is between the masses (or radii) of Earth and Neptune, and the correlation under discussion applies specifically to the super Earths from the inner region. These super Earths dominate the known inner planet population, and they can coexist with almost all types of inner planets except hot Jupiters (see Section 2.1). For this reason, this super Earth population is representative of the inner planet population.

About one-third of the inner super Earths have outer cold Jupiter companions, as studies have shown (Zhu & Wu 2018, Bryan et al. 2019). The RV signal on the star induced by a super Earth is systematically smaller than the RV signal induced by a cold Jupiter. Making use of this point, Zhu & Wu (2018) constructed a sample of 54 super Earth systems around Sun-like hosts that received long-term RV observations, and they found that the fraction of super Earth hosts with cold Jupiter companions is  $32 \pm 8\%$ . This is about three times higher than the frequency of cold Jupiters around field Sun-like stars. The fraction further rises to  $\sim 60\%$  for metal-rich systems (with  $[\text{Fe}/\text{H}] > 0.1$ ). These results were later confirmed by the independent study by Bryan et al. (2019). In that work, the authors refit RV data sets of 65 super Earth hosts, some of which are M dwarfs, and reported an occurrence rate of  $39 \pm 7\%$  for companions with masses in the range of  $0.5\text{--}20 M_J$  and semimajor axes in the range of  $1\text{--}20$  AU. In this review, we take a rather conservative value of  $P(\text{CJ}|\text{SE}) \approx 30\%$ , which is also shown in **Figure 9**.

The inversion of the above conditional probability reveals an even more interesting result. With  $\sim 30\%$  of Sun-like stars hosting inner super Earths and  $\sim 10\%$  of Sun-like stars hosting cold Jupiters, one finds from the Bayes theorem that  $P(\text{SE}|\text{CJ}) \approx 90\%$ , suggesting that nearly all of the cold Jupiters should have inner small planets (Zhu & Wu 2018, Bryan et al. 2019). Together with the fraction of cold Jupiter hosts with hot Jupiters,  $P(\text{HJ}|\text{CJ}) \approx 7\%$ , outer giant planets almost all have inner companions.<sup>11</sup> We illustrate in **Figure 9** the connections between the outer giant planets and the two representative types of inner planets.

The above strong correlations are also confirmed by studies that utilized the rare but valuable long-period *Kepler* transiting planets (Uehara et al. 2016, Herman et al. 2019, Masuda et al. 2020). These studies find that the fraction of long-period ( $P \gtrsim 2$  years) transiting planets with inner transiting companions is so high that it can only be explained by a strong inner–outer correlation. They also reported evidence that the dynamical hotness of inner and outer planets may also be

<sup>11</sup>It is possible that hot Jupiters were born cold and that their later evolution cleared out the small planets originally present in the inner region. This would mean that essentially all cold Jupiters were born with inner small planets.

correlated. Specifically, a dynamically hot outer Jupiter is likely associated with a dynamically hot inner planetary system. This provides a plausible explanation for the surprisingly large eccentricities and mutual inclinations of the inner systems with low multiplicities (Masuda et al. 2020; see Section 2.3). It may also help explain the reduced super Earth multiplicities around metal-rich stars (Zhu & Wu 2018; see Section 2.5.2).

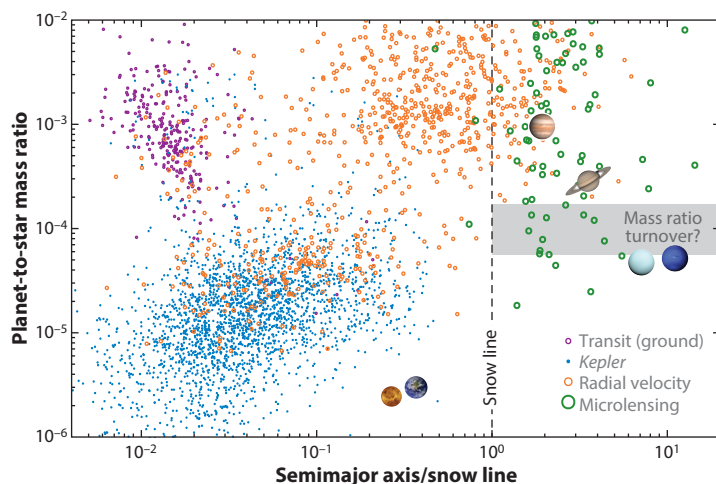
The strong inner–outer correlation has implications on the frequency of planetary systems similar to our own. On the one hand, this correlation suggests that the general Solar System–like architecture—with the inner region containing small planets and the outer region containing giant planets—is probably common among other planetary systems. On the other hand, planetary systems with properties very similar to ours, namely a system with both outer Jupiter-like ( $\sim 1 M_J$  at a few astronomical units) and inner Earth-like ( $\lesssim 1 M_\oplus$  within 1 AU) planets, may be rare ( $\lesssim 1\%$ ; Zhu & Wu 2018). A possible explanation could be that our Jupiter formed very early and, hence, prevented the growth of inner embryos into super Earths (Izidoro et al. 2015, Morbidelli et al. 2015). This early Jupiter formation scenario also explains the isotope measurements on Solar System iron meteorites (Kruijer et al. 2017), although the question remains why the majority of cold Jupiters in other systems do have inner super Earths. We defer further discussions of theoretical implications to Section 4.2.

Observationally, the strong inner–outer correlation implies interesting synergies between space-based transit missions and astrometric missions or ground-based long-term RV surveys. Indeed, at least two of the TESS transiting planets have been found around stars with known RV cold Jupiters (Huang et al. 2018, Teske et al. 2020). Future combined TESS and *Gaia* planet catalogs should yield hundreds of similar systems that can enable detailed studies of the system architecture, as has been demonstrated in the  $\pi$  Mensae system (Damasso et al. 2020, De Rosa et al. 2020, Xuan & Wyatt 2020).

### 3.3. Mass Ratio Function from Microlensing

Gravitational microlensing probes a largely uncharted planet discovery space of cold planets (Mao & Paczynski 1991, Gould & Loeb 1992), where  $>99\%$  of the planetary mass of the Solar System resides. Ground-based microlensing surveys are sensitive to planets down to Earth masses (e.g., Bennett & Rhie 1996, Dong et al. 2006), and a space-based survey will be capable of discovering all Solar System planet analogs except Mercury (Penny et al. 2019; see also **Figure 1**). With the increasing number of discoveries, microlensing searches have been continuing to unveil the distribution of planets in this underexplored parameter space and offer insights into the planet formation outside the water snow line. We refer interested readers to Gaudi (2012) for an overview of the microlensing technique and its application in exoplanet discoveries (see also Mao 2012). Below, we focus on the important progress made since the review by Gaudi (2012).

Microlensing samples directly yield planet-to-star mass ratio distributions (e.g., Shvartzvald et al. 2016, Suzuki et al. 2016), and several recent studies reported a possible turnover in the mass ratio function for planets beyond the water snow line (see **Figure 10**). In the early era, a key microlensing finding was that cold Neptunes (with planet-to-star mass ratio  $q \sim 10^{-4}$ ) are a factor of a few more common than cold Jupiters (with  $q \sim 10^{-3}$ ; Gould et al. 2006b), and Sumi et al. (2010) found that the distribution of the mass ratio  $q$  could be described by a power law  $dN/d\log_{10} q \propto q^\nu$ , where  $\nu = -0.7 \pm 0.2$ . Using the planet sample from the second phase of the Microlensing Observations in Astrophysics survey, Suzuki et al. (2016) found that a single power law of the mass ratio function does not extend to very low mass ratios. Specifically, these authors reported a break in the mass ratio function at  $q \sim 10^{-4}$ , corresponding to the mass of Neptune for a typical host star mass of  $0.5 M_\odot$ . Based on a total sample of 30 planets that combines the MOA-II and previous



**Figure 10**

Similar to **Figure 1**, but here we show the planet-to-star mass ratio versus the semimajor axis in units of the water snow line. The snow line is at 2.7 AU for a  $1\text{-}M_{\odot}$  star and scales linearly with the host mass (Kennedy & Kenyon 2008). The location of the snow line is indicated with a black dashed line (note that this is only for illustrative purposes as the snow line should be determined in the protoplanetary disk), and the possible turnover in the mass-ratio distribution found from microlensing ( $0.55\text{--}1.7 \times 10^{-4}$ ; Suzuki et al. 2016, Jung et al. 2019) is marked with the gray region. Six Solar System planets are shown, with Mercury and Mars being too low in mass ratio to appear on this plot.

statistical samples (Gould et al. 2010, Cassan et al. 2012), Suzuki et al. (2016) reported a broken power-law mass ratio function with a break at  $q_{\text{brk}} = 1.7 \times 10^{-4}$ . The power-law indexes above and below the break are  $\nu = -0.93 \pm 0.13$  and  $0.6^{+0.5}_{-0.4}$ , respectively. The normalization is such that  $\bar{n}_{\text{p}} = 0.79$  for planets with mass ratio  $q > 5 \times 10^{-5}$  and projected separation  $s$  in units of Einstein radius in the range of 0.3–5. For typical microlenses with  $0.5 M_{\odot}$ , these numbers correspond to the planetary mass  $M_{\text{p}} > 8 M_{\oplus}$  and the orbital separation of 1–15 AU. The reported planet frequency is compatible with those from other detection techniques (i.e., RV and direct imaging) following a simple joint planet distribution function (Clanton & Gaudi 2014, 2016).

Further studies by Udalski et al. (2018), who studied an ensemble of 7 (as compared to 4 in Suzuki et al. 2016) planets with  $q < 10^{-4}$ , and Jung et al. (2019), who analyzed a sample of 15 planets with  $q < 3 \times 10^{-4}$ , investigated the possible turnover in the mass ratio function. Adopting a power-law form of the detection efficiency, Jung et al. (2019) modeled the intrinsic mass-ratio distribution with a broken power law and revised the break to  $q_{\text{brk}} \approx 5.5 \times 10^{-5}$ , which is a factor of three below the value found by Suzuki et al. (2016), but their low-mass planet sample was too small to distinguish a pile-up at that mass ratio from broken power law. Nevertheless, a break or pile-up in the planet-to-star mass ratio function could have important theoretical implications (e.g., Pascucci et al. 2018, Wu 2019), and further probing the distribution of sub-Neptune microlensing planets will be a research focus in the near future. Observations from high-cadence and nearly continuous microlensing surveys such as KMTNet (Kim et al. 2016) are pushing toward detecting more planets at low mass ratios (e.g.,  $q = 1.8 \times 10^{-5}$  from Gould et al. 2020a,  $q \approx 1.4 \times 10^{-5}$  from Yee et al. 2021, and  $q \approx 1.1 \times 10^{-5}$  from Zang et al. 2021), so a large enough sample is expected to be available soon to improve the determination of the mass ratio function at the low end (see Zang et al. 2021 and discussions therein).

Another interesting feature of the mass ratio function of Suzuki et al. (2016) is its apparent smoothness between Neptune and Jupiter masses. Intriguingly, the derived radius distribution of cold planets from the single transit events in *Kepler* also appears similarly continuous between Neptune and Saturn (Herman et al. 2019). These results are surprising in view of the standard core accretion theory (Pollack et al. 1996), which builds on the Solar System and predicts a deficit of planets at such intermediate masses and/or radii (Ida & Lin 2004a, Mordasini et al. 2009). This tension may suggest that the giant planet formation involves physical processes that have been overlooked in the standard models (Suzuki et al. 2018). Alternatively, it could be due to the limited sample sizes of cold intermediate-mass planets (the Suzuki et al. 2016 sample contains nine detections in the range of  $10^{-4} < q < 5 \times 10^{-4}$ , and the Herman et al. 2019 sample has four in the intermediate radius bin of  $0.67\text{--}1.00 R_J$ ). Therefore, increasing the sample size of cold intermediate-mass planets will clarify the degree of tension between observation and theory. Furthermore, physical mass (rather than mass ratio) determinations of a large sample of microlensing planets through measurements of the lens flux or the microlensing parallax (e.g., Dong et al. 2009, Udalski et al. 2015b, Yee et al. 2015) are needed to enable a more direct comparison with theories. This will be possible for essentially all microlensing planets detected to date at first light of adaptive optics on 30-m-class telescopes (e.g., Skidmore et al. 2015) or for a significant fraction of planet hosts in a space-based microlensing survey such as the microlensing survey planned for the *Nancy Grace Roman Space Telescope* (Penny et al. 2019).

### 3.4. Free-Floating Planets

The prevalence of eccentric and/or inclined planetary orbits suggests likely histories of violent dynamical interactions in the planetary systems, such as planet–planet scatterings, which naturally eject a significant fraction of the planets from the system and form unbound planets with no hosts (e.g., Rasio & Ford 1996, Chatterjee et al. 2008, Jurić & Tremaine 2008). The distributions of FFPs bear important signatures of not only the initial configurations of the planetary systems at birth but also their subsequent dynamical evolution.

Although it is possible to directly image young substellar objects down to a few Jupiter masses (see, e.g., Zapatero Osorio et al. 2000), gravitational microlensing is the only known method in probing the lower-mass objects, which are believed to dominate the dynamically ejected FFP population. Low-mass objects produce relatively short-timescale microlensing light curves as the Einstein radius crossing time  $t_E \propto \sqrt{M}$ . For typical stellar-mass microlenses, the timescale is  $\sim 20$  days, whereas for planetary-mass objects it is  $\lesssim 1$  day. The detection of such short and rare events thus demands wide-field high-cadence surveys that have only been available since the past decade.

Mróz et al. (2017b) analyzed a sample of 2,617 microlensing events from the OGLE-IV (fourth phase of the Optical Gravitational Lensing Experiment; Udalski et al. 2015a) survey and concluded that the frequency of Jupiter-mass free-floating (or wide-orbit) planets should be no more than 0.25 planets per main-sequence star at the 95% confidence level. This result is broadly compatible with the inferred frequency of bound giant planets from RV surveys (e.g., Cumming et al. 2008), microlensing searches (e.g., Gould et al. 2010), or direct imaging (e.g., Bowler 2016), and contradicts a previous claim that free-floating Jupiter-mass planets are more abundant than stars (Sumi et al. 2011). The sample of Mróz et al. (2017b) also includes six short events with timescales in the range of  $0.1 \text{ days} < t_E < 0.4 \text{ days}$ . Assuming the microlensing nature of these events and given the low detection efficiency at such ultrashort timescales, their sample suggests that there may be up to a few FFPs in the Earth-mass to super-Earth-mass range per main-sequence star. The results of Mróz et al. (2017b) about the absence of free-floating Jupiter-mass planets and the potential

**Table 1** Published microlensing free-floating planet candidates ( $\theta_E < 10 \mu\text{as}$ ), sorted by the inferred lens mass (see Equations 22 and 23 for typical estimates in the bulge and disk, respectively)

Event name	$\theta_E/\mu\text{as}$	$t_E/\text{d}$	$M(\text{disk})$	$M(\text{bulge})$	Reference
OGLE-2016-BLG-1928	0.84	0.029	$0.2 M_\oplus$	$1.8 M_\oplus$	Mróz et al. 2020a
OGLE-2012-BLG-1323	2.4	0.16	$1.8 M_\oplus$	$14 M_\oplus$	Mróz et al. 2019
OGLE-2019-BLG-0551	4.4	0.38	$6.1 M_\oplus$	$48 M_\oplus$	Mróz et al. 2020b
KMT-2019-BLG-2073	4.8	0.27	$7.6 M_\oplus$	$59 M_\oplus$	Kim et al. 2020
KMT-2017-BLG-2820	5.9	0.29	$11 M_\oplus$	$87 M_\oplus$	Ryu et al. 2021
OGLE-2016-BLG-1540	9.2	0.32	$28 M_\oplus$	$217 M_\oplus$	Mróz et al. 2018

existence of free-floating Earth-mass and super-Earth-mass planets are generally consistent with theoretical expectations (e.g., Ida et al. 2013, Ma et al. 2016).

The existence of such ultrashort-timescale events was soon confirmed thanks to the coordinated observations of multiple microlensing survey telescopes around the globe. Mróz et al. (2018) reported the first convincing example of a microlensing event with timescale  $t_E = 0.32$  days, and subsequent dedicated searches led to the discovery of a few more similar events (Kim et al. 2020; Mróz et al. 2019, 2020a,b; Ryu et al. 2021). These events all show strong finite-source effects that arise from the lenses transiting distant giant sources, yielding the immediate measurement of the angular Einstein radius  $\theta_E$ . The lens mass scales as

$$M(\text{bulge}) = \frac{\theta_E^2}{\kappa \pi_{\text{rel}}} = 250 M_\oplus \left( \frac{\theta_E}{10 \mu\text{as}} \right)^2 \left( \frac{\pi_{\text{rel}}}{16 \mu\text{as}} \right)^{-1}, \quad 22.$$

with the normalization of the lens-source relative parallax  $\pi_{\text{rel}}$  chosen such that the lens and source are both in the Galactic bulge and separated by about 1 kpc in front. Here, the constant  $\kappa \approx 8.14 \text{ mas M}^{-1}$ . For lens in the Galactic disk ( $\pi_{\text{rel}} \approx 125 \mu\text{as}$ ), the mass scales as

$$M(\text{disk}) = \frac{\theta_E^2}{\kappa \pi_{\text{rel}}} = 32 M_\oplus \left( \frac{\theta_E}{10 \mu\text{as}} \right)^2 \left( \frac{\pi_{\text{rel}}}{125 \mu\text{as}} \right)^{-1}. \quad 23.$$

Kim et al. (2020) and Ryu et al. (2021) argue that  $\theta_E$  is a better discriminator than  $t_E$  for selecting FFPs. In fact, from a small number of events with finite-source effects, there is a possible gap between  $\sim 10 \mu\text{as}$  and  $\sim 30 \mu\text{as}$  in the  $\theta_E$  distribution, and this Einstein desert may separate brown dwarfs from free-floating super Earths (and terrestrial planets) in the disk (Ryu et al. 2021). We list in **Table 1** the relevant parameters and inferred masses of the FFP candidate events with  $\theta_E < 10 \mu\text{as}$ . The preliminary analyses by Mróz et al. (2019) and Ryu et al. (2021) suggest that low-mass unbound (or wide-orbit) planets are possibly more common than stars in the Galaxy. Future space-based microlensing surveys can assemble a large sample for quantitative assessments of the FFP population (e.g., Johnson et al. 2020), and a satellite augmented with microlensing parallax measurements can directly measure the masses and distances of such FFP events (e.g., Gould et al. 2020b, and references therein).

Although the events listed in **Table 1** are promising candidates for FFPs, it is also plausible that these objects are actually in such wide orbits that no microlensing signatures from their hosts were detected. Light curve analyses can exclude the existence of any massive companions (i.e., hosts) out to a few Einstein radii, corresponding to  $\sim 15\text{--}20$  AU away (e.g., Mróz et al. 2018, Kim et al. 2020). In other words, these FFP candidates could well be planets at Uranus-like or Neptune-like orbits (e.g., Poleski et al. 2014). Future high-resolution imaging observations that can resolve the hosts for wide-separation planets will be able to tell whether these objects are truly

free-floating or loosely bound to some unidentified stellar hosts (Han et al. 2005, Gould 2016, Ryu et al. 2021).

## 4. THEORETICAL IMPLICATIONS

The observed distribution of planets and the architecture of planetary systems, as reviewed in Sections 2 and 3, are the consequences of  $\sim 10$ – $100$ -Myr formation and later  $\sim 1$ -Gyr evolution. In this section, we discuss the constraints from these observations on theoretical models. A comprehensive overview on the formation and evolution theories is beyond the scope of the current review. Instead, we focus on the key physical processes that lead to observational signatures. To reduce the complexity, we restrict our discussion to planetary systems around Sun-like hosts.

### 4.1. A Brief Overview of Theories

The generally accepted picture of planet formation can be traced back to the nebular model originally proposed by Immanuel Kant and Pierre Laplace in the 1700s. Modern theorists generally believe that planets were formed out of the gas and the dust in the protoplanetary disk. Small solid particles first accumulate to form asteroid-sized ( $\sim 1$ – $100$  km) planetesimals, and the collisions between planetesimals eventually lead to the formation of planet-sized objects (Chamberlin 1916, Safronov 1972). See Woolfson (1993) for a historical overview on the planet-formation theories.

In the core accretion theory that explains our Solar System's formation (e.g., Lissauer 1993, Pollack et al. 1996), the primary building blocks for planet formation are planetesimals. The growth of planetesimals is first divergent (i.e., the runaway phase) and then convergent (i.e., the oligarchic phase), until nearly all planetesimals in their feeding zones are cleared ( $\sim 5 R_H$ ). These so-called protoplanets (or embryos) are now  $\gtrsim 1,000$  km in size and around Mars-mass (e.g., Ida & Makino 1993, Kokubo & Ida 1998). The further growth of the protoplanets involves planetesimal accretion as well as dynamical interactions between protoplanets. At a few astronomical units separation, the growth of protoplanets is sufficient and allows the formation of giant planets (Mizuno 1980, Pollack et al. 1996). In the classical picture, the giant planet formation has three phases: core formation, hydrostatic gas accretion, and runaway gas accretion (Pollack et al. 1996). The hydrostatic gas accretion phase starts when the embryo reaches a critical core mass ( $\sim 10 M_\oplus$ ; Mizuno 1980, Stevenson 1982). This phase can take up to  $\sim 10$  Myr and is the most time-consuming step in this classical core accretion model. The runaway gas accretion is triggered once the envelope and the core have comparable masses, and it sufficiently pushes the total mass to the giant planet regime ( $\gtrsim 100 M_\oplus$ ). In the inner region, embryos grow slowly and never reach the critical core mass before the gaseous disk is depleted. The later evolution involves collisions between these embryos in the gas-free environment. This so-called giant impact phase lasts  $\sim 100$  Myr and eventually forms the terrestrial planets (e.g., Chambers 2001).

A new paradigm that has attracted much attention in recent years is pebble accretion. In the astrophysical context, pebbles are dust particles that are weakly coupled to the gas and, thus, drift in the disk. The inclusion of pebbles in the formation diagram provides a plausible scenario for the formation of planetesimals via streaming instability (Youdin & Goodman 2005, Johansen et al. 2007, Chiang & Youdin 2010). Unlike planetesimals that are decoupled from the gas, pebbles “feel” the aerodynamic drag from the gas and drift inward toward the star (Nakagawa et al. 1986). This means that the food supply to a protoplanet is not limited to the local material. Additionally, the cross-section for protoplanets (or planetesimals) to accrete pebbles is larger than the cross-section for the same objects to accrete planetesimals (Ormel & Klahr 2010, Lambrechts & Johansen 2012). These two factors together make pebble accretion more efficient in building up cores of protoplanets. When the protoplanet becomes massive enough, it starts to carve a gap in



the pebble disk, and the subsequent pressure bump outside of the orbit stops the inward drifting pebbles. The corresponding mass of the protoplanet is called the pebble isolation mass,

$$M_{\text{iso}} \approx 10 \left( \frac{b/r}{0.04} \right)^3 M_{\oplus}, \quad 24.$$

where  $b/r$  is the disk aspect ratio at location  $r$  and the prefactor is determined numerically and depends on disk properties (Lambrechts et al. 2014). The above relation assumes a solar-mass host star. For other stellar masses, the pebble isolation mass scales linearly with the mass of the host star (Liu et al. 2019). Once a protoplanet reaches the pebble isolation mass, it effectively cuts off the pebble flux and starves the protoplanetary core and all embryos interior to its orbit. With the halted pebble accretion, the critical core mass required to trigger the rapid gas accretion is reduced and thus giant planets can form more efficiently (Lambrechts et al. 2014). Furthermore, pebbles can easily vaporize in the hot envelopes before they can reach the cores (Brouwers et al. 2018). The enriched envelopes also speed up the formation of giant planets (Venturini et al. 2016; see also Stevenson 1982 and Hori & Ikoma 2011 for a similar mechanism in the planetesimal accretion scenario). We refer interested readers to the reviews by Johansen & Lambrechts (2017) and Ormel (2017) for more details about the pebble accretion model.

Planets may undergo disk-driven migration while accreting pebbles, planetesimals, and/or gas (e.g., Kley & Nelson 2012, and references therein). Migration can substantially change the architecture of the planetary system, such as locking planets into mean motion resonances (Goldreich & Tremaine 1980, Lee & Peale 2002). However, such features are not prominent in *Kepler* systems (see Section 2.4.2). This may suggest that most *Kepler* planets have not undergone significant disk-driven migrations (see also Section 4.2). Alternatively, the *Kepler* planets may have never entered into resonances during the migration (e.g., Goldreich & Schlichting 2014), or the long-term dynamical evolution after the disk dispersal has effectively removed most of these features (e.g., Izidoro et al. 2017, 2019).

## 4.2. Constraints from Observations

Given the substantial uncertainties in theories and in some parts of observations, we think that it is premature to provide detailed and quantitative comparisons between theories and observations (but see attempts by, e.g., Hansen & Murray 2013, Izidoro et al. 2017, Mulders et al. 2019, Emsenhuber et al. 2020). We therefore choose to focus on the following selected constraints that are considered relatively robust and discuss their implications to the formation of *Kepler*-like planets:

- **Prevalence and multiplicity.** Inner super Earth-like planets are known to exist around  $\sim 30\%$  of Sun-like stars, and they typically reside in multiplanet systems (Section 2.2). Additionally, they preferentially have outer cold Jupiter companions (Section 3.2), suggesting that the two types of planets do not inhibit, but perhaps promote, the formation of each other. Unlike giant planets, super Earths show a much weaker dependence on the host metallicity (Section 2.5.2).
- **Composition.** As inferred from population-level studies of the radius valley (Section 2.1.4) as well as mass and radius measurements of individual *Kepler* planets (e.g., Wu & Lithwick 2013; Hadden & Lithwick 2014, 2017), some inner small planets likely have Earth-like (i.e., rocky and ice-poor) cores, and these cores have acquired gaseous envelopes that weigh up to a few percent of the total mass while the disk is still present.



The prevalence and the early formation of super Earth–like planets suggest that the planet-formation process is more efficient than what had been expected from Solar System formation models. This alone may not be an issue to the pebble accretion scenario (see Section 4.1).<sup>12</sup> In fact, pebble accretion can be so efficient that preventing super Earth–mass planets from undergoing runaway gas accretion poses another challenge, a possible solution to which could be a delayed formation near the end of the disk phase (e.g., Lee et al. 2014). For the planetesimal accretion scenario, a very massive disk is typically required to form super Earths efficiently and early (e.g., Emsenhuber et al. 2020). The rocky composition suggests that the cores are formed in the ice-poor environment, likely inside the water ice line. In order for embryos or protoplanets from outside of the ice line to not largely contaminate the inner region, the disk-driven migration is probably suppressed.

The strong correlation between inner super Earths and outer cold Jupiters is a bit challenging to both accretion scenarios under the typical protoplanetary disk conditions. The planetesimal accretion scenario usually requires relatively efficient disk-driven migrations to explain the presence of abundant super Earths around metal-poor hosts, but the same migration efficiency turns out to be an overkill in reproducing the inner–outer strong correlation (Schlecker et al. 2020; see also Ida & Lin 2010). For the pebble accretion scenario, because the solid supply to protoplanets is not limited locally, there is potentially direct competition between different embryos. Furthermore, once the core of the outer giant planet first reaches the pebble isolation mass, the further growth of the inner planets is significantly limited, and the giant planet also acts as a barrier to the inward migrating embryos from outside of its orbit. Therefore, the pebble accretion scenario typically expects an anticorrelation between inner and outer planets (Izidoro et al. 2015, Morbidelli et al. 2015, Lambrechts et al. 2019). Alternatively, the cores of both inner super Earths and outer cold Jupiters could be formed at such large separations (tens of astronomical units) that enough material is available to the inner cores (Bitsch et al. 2015, 2019), although it is unclear whether such an approach can reproduce quantitatively the observed correlation and form rocky core planets. The difficulty in reproducing the inner–outer correlation may suggest that many protoplanetary disks start heavier than what has been typically assumed. Indeed, if *Kepler* planets are formed in situ based on the local material (i.e., not the inward-drifting pebbles), the required surface density is much higher than in the minimum-mass solar nebula model (Weidenschilling 1977, Hayashi 1981) and almost reaches the gravitational instability limit (Chiang & Laughlin 2013).

## 5. SUMMARY AND DISCUSSION

The discovery of thousands of exoplanets from the combination of multiple detection techniques has substantially advanced our understanding of the distribution of planets and the architecture of planetary systems. This review aims to update our knowledge of exoplanet statistics since the Winn & Fabrycky (2015) review. In Section 2, we described the distribution and properties of planets in the inner region, based mostly on discoveries from the *Kepler* mission. In Section 3, we reviewed the recent progress on the cold planet population, with an emphasis on their connections to the close-in companions. Section 4 briefly described the theoretical models and the key constraints from observations. With the ongoing and upcoming missions that have better capabilities and/or open up new observational channels, our understanding of exoplanets and planetary

<sup>12</sup> Pebble accretion is efficient in growing planet embryos into larger bodies. However, pebble accretion is also lossy, as  $\gtrsim 90\%$  of the planet-forming material falls onto the host star rather than being accreted onto the growing planets (Liu & Ormel 2018, Lin et al. 2018).

systems will continue to be improved. Below we summarize the key results and outline several promising directions that may see substantial advancement in the near future.

## SUMMARY POINTS

1. In the inner region ( $\lesssim 1$  AU), about 30% of Sun-like stars host planets with masses and/or radii down to Earth mass and/or radius, and each planetary system on average has about three such planets. These suggest that the planet-formation process is more efficient than what had been expected from Solar System formation models.
2. Planetary systems with more planets appear colder dynamically, with smaller orbital eccentricities, mutual inclinations, and orbital spacings. For systems with few planets in the inner region, planets can have  $\sim 0.3$  orbital eccentricities and  $\gtrsim 10$ -deg mutual inclinations. These support the idea that dynamical evolution has played a significant role reshaping the system architecture.
3. There exists a radius valley at  $R_p \sim 2 R_\oplus$  and  $P \lesssim 30$  days. The valley was predicted by the photoevaporation theory, although alternative explanations have also been proposed. Population-level analyses of the radius valley suggest that these planets were probably born with rocky cores and gaseous atmospheres up to a few percent of the core masses.
4. Cold Neptune-like planets are a few times more abundant than cold Jupiter-like ones in the outer region. The inner ( $\lesssim 1$  AU) and the outer ( $\sim 1$ – $10$  AU) planetary systems appear strongly correlated such that inner small planets preferentially have cold Jupiter-like companions and outer Jupiters almost always have inner planetary companions.

## FUTURE ISSUES

1. Exoplanet atmosphere and mass–radius relation: Space-based all-sky transit surveys like the TESS (*Transiting Exoplanet Survey Satellite*) mission (Ricker et al. 2015) have been finding many bright targets, enabling detailed characterizations of more close-in planets (e.g., Huang et al. 2018, Armstrong et al. 2020). An improved mass–radius relation and better atmospheric characterizations will help to understand the composition and potentially the past evolution of the planet (see the recent review by Madhusudhan 2019).
2. Planetary system architecture: The joint coverage of different surveys will potentially open up a larger parameter space (see **Figure 1**) and reveal more interesting features about the planetary system architecture. The *Gaia* mission alone is expected to detect at least thousands of giant planets around nearby stars (Perryman et al. 2014), and its synergy with other surveys and missions will also open up new channels into architecture study (e.g., Xuan & Wyatt 2020, Damasso et al. 2020, De Rosa et al. 2020). Radial velocity (RV) follow-ups of systems detected by other methods could also play increasingly indispensable roles in this aspect, particularly with the extreme RV instruments with capability down to  $\sim 0.3 \text{ m s}^{-1}$  now coming online (e.g., Fischer et al. 2016).
3. Planets across the Hertzsprung–Russell diagram and their Galactic distributions: Rapid advances of large-scale spectroscopic surveys and the *Gaia* satellite are continuing to revolutionize stellar astrophysics and Galactic astronomy. Further synergies of large samples of exoplanets with detailed stellar chemical compositions, kinematics, and/or

ages measurements are expected to place planet formation in the rich context of stellar populations and evolutions.

4. Planets around young stars: Although the present review focuses on the planetary systems around  $\gtrsim 1$ -Gyr-old stars, the demographics of planets around young ( $\lesssim 100$  Myr) stars is a crucial link toward a more direct comparison with both planet formation theories and ALMA (Atacama Large Millimeter/submillimeter Array) observations of protoplanetary disks (see the recent review by Andrews 2020). The detection and characterization of more planets around young stars [e.g., PDS 70b, c (Haffert et al. 2019) and AU Microscopii b (Plavchan et al. 2020)] will be valuable.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

We thank Scott Gaudi, Andy Gould, Kento Masuda, Shude Mao, Chris Ormel, Scott Tremaine, and Josh Winn for comments and suggestions on the manuscript. S.D. was supported by the National Key R&D Program of China (grant no. 2019YFA0405100). W.Z. was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the funding reference #CITA 490888-16.

## LITERATURE CITED

- Adibekyan VZ, Sousa SG, Santos NC, et al. 2012. *Astron. Astrophys.* 545:A32
- Agol E, Steffen J, Sari R, Clarkson W. 2005. *MNRAS* 359:567–79
- Akeson RL, Chen X, Ciardi D, et al. 2013. *Publ. Astron. Soc. Pac.* 125:989
- Andrews SM. 2020. *Annu. Rev. Astron. Astrophys.* 58:483–528
- Andrews SM, Rosenfeld KA, Kraus AL, Wilner DJ. 2013. *Ap. J.* 771:129
- Ansdell M, Williams JP, van der Marel N, et al. 2016. *Ap. J.* 828:46
- Armstrong DJ, Lopez TA, Adibekyan V, et al. 2020. *Nature* 583:39–42
- Artymowicz P, Lubow SH. 1994. *Ap. J.* 421:651–67
- Bailey E, Batygin K. 2018. *Ap. J. Lett.* 866:L2
- Ballard S, Johnson JA. 2016. *Ap. J.* 816:66
- Beaugé C, Nesvorný D. 2013. *Ap. J.* 763:12
- Becker JC, Vanderburg A, Adams FC, Rappaport SA, Schwengeler HM. 2015. *Ap. J. Lett.* 812:L18
- Bennett DP, Rhie SH. 1996. *Ap. J.* 472:660–64
- Berger TA, Huber D, Gaidos E, van Saders JL. 2018. *Ap. J.* 866:99
- Berger TA, Huber D, Gaidos E, van Saders JL, Weiss LM. 2020a. *Astron. J.* 160:108
- Berger TA, Huber D, van Saders JL, et al. 2020b. *Astron. J.* 159:280
- Bitsch B, Izidoro A, Johansen A, et al. 2019. *Astron. Astrophys.* 623:A88
- Bitsch B, Lambrechts M, Johansen A. 2015. *Astron. Astrophys.* 582:A112
- Borucki WJ, Koch D, Basri G, et al. 2010. *Science* 327:977
- Bouma LG, Masuda K, Winn JN. 2018. *Astron. J.* 155:244
- Bowler BP. 2016. *Publ. Astron. Soc. Pac.* 128:102001
- Brewer JM, Wang S, Fischer DA, Foreman-Mackey D. 2018. *Ap. J. Lett.* 867:L3
- Brouwers MG, Vazan A, Ormel CW. 2018. *Astron. Astrophys.* 611:A65

- Bryan ML, Knutson HA, Howard AW, et al. 2016. *Ap. J.* 821:89
- Bryan ML, Knutson HA, Lee EJ, et al. 2019. *Astron. J.* 157:52
- Buchhave LA, Bitsch B, Johansen A, et al. 2018. *Ap. J.* 856:37
- Buchhave LA, Latham DW, Johansen A, et al. 2012. *Nature* 486:375–77
- Burke CJ, Catanzarite J. 2017a. Planet Detection Metrics: Per-Target Detection Contours for Data Release 25. Kepler Sci. Doc. KSCI-19111-002
- Burke CJ, Catanzarite J. 2017b. Planet Detection Metrics: Per-Target Flux-Level Transit Injection Tests of TPS for Data Release 25. Kepler Sci. Doc. KSCI-19109-002
- Burke CJ, Christiansen JL, Mullally F, et al. 2015. *Ap. J.* 809:8
- Burt JA, Nielsen LD, Quinn SN, et al. 2020. *Astron. J.* 160:153
- Cañas CI, Wang S, Mahadevan S, et al. 2019. *Ap. J. Lett.* 870:L17
- Cassan A, Kubas D, Beaulieu JP, et al. 2012. *Nature* 481:167–69
- Chamberlin TC. 1916. *J. R. Astron. Soc. Can.* 10:473
- Chambers JE. 2001. *Icarus* 152:205–24
- Chambers JE, Wetherill GW, Boss AP. 1996. *Icarus* 119:261–68
- Chaplin WJ, Miglio A. 2013. *Annu. Rev. Astron. Astrophys.* 51:353–92
- Chatterjee S, Ford EB, Matsumura S, Rasio FA. 2008. *Ap. J.* 686:580–602
- Chen J, Kipping D. 2017. *Ap. J.* 834:17
- Chiang E, Laughlin G. 2013. *MNRAS* 431:3444–55
- Chiang E, Youdin AN. 2010. *Annu. Rev. Earth Planet. Sci.* 38:493–522
- Christiansen JL, Clarke BD, Burke CJ, et al. 2020. *Astron. J.* 160:159
- Christiansen JL, Clarke BD, Burke CJ, et al. 2015. *Ap. J.* 810:95
- Ciardi DR, Beichman CA, Horch EP, Howell SB. 2015. *Ap. J.* 805:16
- Ciardi DR, Fabrycky DC, Ford EB, et al. 2013. *Ap. J.* 763:41
- Clanton C, Gaudi BS. 2014. *Ap. J.* 791:91
- Clanton C, Gaudi BS. 2016. *Ap. J.* 819:125
- Cui X-Q, Zhao Y-H, Chu Y-Q, et al. 2012. *Res. Astron. Astrophys.* 12:1197–242
- Cumming A, Butler RP, Marcy GW, et al. 2008. *Publ. Astron. Soc. Pac.* 120:531
- Dai F, Masuda K, Winn JN. 2018. *Ap. J. Lett.* 864:L38
- Damasso M, Sozzetti A, Lovis C, et al. 2020. *Astron. Astrophys.* 642:A31
- Dawson RI, Chiang E. 2014. *Science* 346:212–16
- Dawson RI, Johnson JA. 2018. *Annu. Rev. Astron. Astrophys.* 56:175–221
- Dawson RI, Murray-Clay RA. 2013. *Ap. J. Lett.* 767:L24
- De Cat P, Fu JN, Ren AB, et al. 2015. *Ap. J. Suppl.* 220:19
- De Rosa RJ, Dawson R, Nielsen EL. 2020. *Astron. Astrophys.* 640:A73
- Deck KM, Payne M, Holman MJ. 2013. *Ap. J.* 774:129
- Dong S, DePoy DL, Gaudi BS, et al. 2006. *Ap. J.* 642:842–60
- Dong S, Gould A, Udalski A, et al. 2009. *Ap. J.* 695:970–87
- Dong S, Katz B, Socrates A. 2014a. *Ap. J. Lett.* 781:L5
- Dong S, Xie JW, Zhou JL, Zheng Z, Luo A. 2018. *PNAS* 115:266–71
- Dong S, Zheng Z, Zhu Z, et al. 2014b. *Ap. J. Lett.* 789:L3
- Dong S, Zhu Z. 2013. *Ap. J.* 778:53
- Dressing CD, Charbonneau D. 2013. *Ap. J.* 767:95
- Dressing CD, Charbonneau D. 2015. *Ap. J.* 807:45
- Duchêne G, Kraus A. 2013. *Annu. Rev. Astron. Astrophys.* 51:269–310
- Emsenhuber A, Mordasini C, Burn R, et al. 2020. *Astron. Astrophys.* Submitted. arXiv:2007.05562
- Fabrycky DC, Lissauer JJ, Ragozzine D, et al. 2014. *Ap. J.* 790:146
- Fang J, Margot JL. 2012. *Ap. J.* 761:92
- Fang J, Margot JL. 2013. *Ap. J.* 767:115
- Fernandes RB, Mulders GD, Pascucci I, Mordasini C, Emsenhuber A. 2019. *Ap. J.* 874:81
- Figueira P, Marmier M, Boué G, et al. 2012. *Astron. Astrophys.* 541:A139
- Fischer DA, Anglada-Escudé G, Arriagada P, et al. 2016. *Publ. Astron. Soc. Pac.* 128:066001

- Fischer DA, Valenti J. 2005. *Ap. J.* 622:1102–17
- Ford EB, Quinn SN, Veras D. 2008. *Ap. J.* 678:1407–18
- Ford EB, Rowe JF, Fabrycky DC, et al. 2011. *Ap. J. Suppl.* 197:2
- Foreman-Mackey D, Hogg DW, Morton TD. 2014. *Ap. J.* 795:64
- Foreman-Mackey D, Morton TD, Hogg DW, Agol E, Schölkopf B. 2016. *Astron. J.* 152:206
- Fressin F, Torres G, Charbonneau D, et al. 2013. *Ap. J.* 766:81
- Fulton BJ, Petigura EA. 2018. *Astron. J.* 156:264
- Fulton BJ, Petigura EA, Howard AW, et al. 2017. *Astron. J.* 154:109
- Funk B, Wuchterl G, Schwarz R, Pilat-Lohinger E, Eggl S. 2010. *Astron. Astrophys.* 516:A82
- Furlan E, Ciardi DR, Everett ME, et al. 2017. *Astron. J.* 153:71
- Gaia Collab., Brown AGA, Vallenari A, et al. 2018. *Astron. Astrophys.* 616:A1
- Gaia Collab., Prusti T, de Bruijne JHJ, et al. 2016. *Astron. Astrophys.* 595:A1
- Gandolfi D, Barragán O, Livingston JH, et al. 2018. *Astron. Astrophys.* 619:L10
- Gaudi BS. 2012. *Annu. Rev. Astron. Astrophys.* 50:411–53
- Gilbert GJ, Fabrycky DC. 2020. *Astron. J.* 159:281
- Ginzburg S, Schlichting HE, Sari R. 2018. *MNRAS* 476:759–65
- Goldreich P, Schlichting HE. 2014. *Astron. J.* 147:32
- Goldreich P, Tremaine S. 1980. *Ap. J.* 241:425–41
- Gould A. 2016. *J. Korean Astron. Soc.* 49:123–26
- Gould A, Dong S, Gaudi BS, et al. 2010. *Ap. J.* 720:1073–89
- Gould A, Dorsher S, Gaudi BS, Udalski A. 2006a. *Acta Astron.* 56:1–50
- Gould A, Loeb A. 1992. *Ap. J.* 396:104–14
- Gould A, Ryu YH, Calchi Novati S, et al. 2020a. *J. Korean Astron. Soc.* 53:9–26
- Gould A, Udalski A, An D, et al. 2006b. *Ap. J. Lett.* 644:L37–40
- Gould A, Zang W, Mao S, Dong S. 2020b. *Res. Astron. Astrophys.* In press. arXiv:2010.09671
- Guo X, Johnson JA, Mann AW, et al. 2017. *Ap. J.* 838:25
- Gupta A, Schlichting HE. 2019. *MNRAS* 487:24–33
- Gupta A, Schlichting HE. 2020. *MNRAS* 493:792–806
- Hadden S, Lithwick Y. 2014. *Ap. J.* 787:80
- Hadden S, Lithwick Y. 2017. *Astron. J.* 154:5
- Hadden S, Lithwick Y. 2018. *Astron. J.* 156:95
- Haffert SY, Bohn AJ, de Boer J, et al. 2019. *Nat. Astron.* 3:749–54
- Hallatt T, Lee EJ. 2020. *Ap. J.* 904:134
- Hamer JH, Schlaufman KC. 2020. *Astron. J.* 160:138
- Han C, Gaudi BS, An JH, Gould A. 2005. *Ap. J.* 618:962–72
- Hansen BMS, Murray N. 2013. *Ap. J.* 775:53
- Hayashi C. 1981. *Prog. Theor. Phys. Suppl.* 70:35–53
- He MY, Ford EB, Ragozzine D. 2019. *MNRAS* 490:4575–605
- He MY, Ford EB, Ragozzine D, Carrera D. 2020. *Astron. J.* 160:276
- Herman MK, Zhu W, Wu Y. 2019. *Astron. J.* 157:248
- Hirano T, Dai F, Gandolfi D, et al. 2018. *Astron. J.* 155:127
- Hoffman KL, Rowe JF. 2017. Uniform Modeling of KOIs: MCMC Notes for Data Release 25. Kepler Sci. Doc. KSCI-19113-001
- Holczer T, Mazeh T, Nachmani G, et al. 2016. *Ap. J. Suppl.* 225:9
- Holman MJ, Murray NW. 2005. *Science* 307:1288–91
- Holman MJ, Wiegert PA. 1999. *Astron. J.* 117:621–28
- Hori Y, Ikoma M. 2011. *MNRAS* 416:1419–29
- Howard AW, Marcy GW, Bryson ST, et al. 2012. *Ap. J. Suppl.* 201:15
- Hsu DC, Ford EB, Ragozzine D, Ashby K. 2019. *Astron. J.* 158:109
- Hsu DC, Ford EB, Ragozzine D, Morehead RC. 2018. *Astron. J.* 155:205
- Huang C, Wu Y, TriAUD AHMJ. 2016. *Ap. J.* 825:98
- Huang CX, Burt J, Vanderburg A, et al. 2018. *Ap. J. Lett.* 868:L39

- Huang CX, Petrovich C, Deibert E. 2017. *Astron. J.* 153:210
- Huang CX, Quinn SN, Vanderburg A, et al. 2020. *Ap. J. Lett.* 892:L7
- Ida S, Kokubo E, Makino J. 1993. *MNRAS* 263:875–89
- Ida S, Lin DNC. 2004a. *Ap. J.* 604:388–413
- Ida S, Lin DNC. 2004b. *Ap. J.* 616:567–72
- Ida S, Lin DNC. 2010. *Ap. J.* 719:810–30
- Ida S, Lin DNC, Nagasawa M. 2013. *Ap. J.* 775:42
- Ida S, Makino J. 1993. *Icarus* 106:210–27
- Izidoro A, Bitsch B, Raymond SN, et al. 2019. *Astron. Astrophys.* Accepted. arXiv:1902.08772
- Izidoro A, Ogihara M, Raymond SN, et al. 2017. *MNRAS* 470:1750–70
- Izidoro A, Raymond SN, Morbidelli A, Hersant F, Pierens A. 2015. *Ap. J. Lett.* 800:L22
- Jackson AP, Davis TA, Wheatley PJ. 2012. *MNRAS* 422:2024–43
- Jackson B, Stark CC, Adams ER, Chambers J, Deming D. 2013. *Ap. J.* 779:165
- Jiang CF, Xie JW, Zhou JL. 2020. *Astron. J.* 160:180
- Jin S, Mordasini C. 2018. *Ap. J.* 853:163
- Jin S, Mordasini C, Parmentier V, et al. 2014. *Ap. J.* 795:65
- Johansen A, Davies MB, Church RP, Holmelin V. 2012. *Ap. J.* 758:39
- Johansen A, Lambrechts M. 2017. *Annu. Rev. Earth Planet. Sci.* 45:359–87
- Johansen A, Oishi JS, Mac Low MM, et al. 2007. *Nature* 448:1022–25
- Johnson JA, Aller KM, Howard AW, Crepp JR. 2010. *Publ. Astron. Soc. Pac.* 122:905–15
- Johnson JA, Butler RP, Marcy GW, et al. 2007. *Ap. J.* 670:833–40
- Johnson JA, Petigura EA, Fulton BJ, et al. 2017. *Astron. J.* 154:108
- Johnson SA, Penny MT, Gaudi BS, et al. 2020. *Astron. J.* 160:123
- Jung YK, Gould A, Zang W, et al. 2019. *Astron. J.* 157:72
- Jurić M, Tremaine S. 2008. *Ap. J.* 686:603–20
- Kane SR, Ciardi DR, Gelino DM, von Braun K. 2012. *MNRAS* 425:757–62
- Kawahara H, Masuda K. 2019. *Astron. J.* 157:218
- Kennedy GM, Kenyon SJ. 2008. *Ap. J.* 673:502–12
- Kim HW, Hwang KH, Gould A, et al. 2020. *Astron. J.* Accepted. arXiv:2007.06870
- Kim SL, Lee CU, Park BG, et al. 2016. *J. Korean Astron. Soc.* 49:37–44
- King GW, Wheatley PJ. 2021. *MNRAS* 501:L28–32
- Kipping D. 2018. *MNRAS* 473:784–95
- Kley W, Nelson RP. 2012. *Annu. Rev. Astron. Astrophys.* 50:211–49
- Knutson HA, Fulton BJ, Montet BT, et al. 2014. *Ap. J.* 785:126
- Kokubo E, Ida S. 1998. *Icarus* 131:171–78
- Kraus AL, Ireland MJ, Huber D, Mann AW, Dupuy TJ. 2016. *Astron. J.* 152:8
- Kruijer TS, Burkhardt C, Budde G, Kleine T. 2017. *PNAS* 114:6712–16
- Kurokawa H, Nakamoto T. 2014. *Ap. J.* 783:54
- Kutra T, Wu Y, Qian Y. 2020. *Astron. J.* Accepted. arXiv:2003.08431
- Lambrechts M, Johansen A. 2012. *Astron. Astrophys.* 544:A32
- Lambrechts M, Johansen A, Morbidelli A. 2014. *Astron. Astrophys.* 572:A35
- Lambrechts M, Morbidelli A, Jacobson SA, et al. 2019. *Astron. Astrophys.* 627:A83
- Laskar J. 1997. *Astron. Astrophys.* 317:L75–78
- Laskar J, Petit AC. 2017. *Astron. Astrophys.* 605:A72
- Lee EJ. 2019. *Ap. J.* 878:36
- Lee EJ, Chiang E. 2017. *Ap. J.* 842:40
- Lee EJ, Chiang E, Ormel CW. 2014. *Ap. J.* 797:95
- Lee EJ, Connors NJ. 2021. *Ap. J.* 908:32
- Lee MH, Peale SJ. 2002. *Ap. J.* 567:596–609
- Limbach MA, Turner EL. 2015. *PNAS* 112:20–24
- Lin DNC, Bodenheimer P, Richardson DC. 1996. *Nature* 380:606–7
- Lin JW, Lee EJ, Chiang E. 2018. *MNRAS* 480:4338–54

- Lissauer JJ. 1993. *Annu. Rev. Astron. Astrophys.* 31:129–74
- Lissauer JJ, Ragozzine D, Fabrycky DC, et al. 2011. *Ap. J. Suppl.* 197:8
- Lithwick Y, Xie J, Wu Y. 2012. *Ap. J.* 761:122
- Liu B, Lambrechts M, Johansen A, Liu F. 2019. *Astron. Astrophys.* 632:A7
- Liu B, Ormel CW. 2018. *Astron. Astrophys.* 615:A138
- Liu F, Yong D, Asplund M, et al. 2016. *MNRAS* 456:2636–46
- Lloyd JP. 2011. *Ap. J. Lett.* 739:L49
- Lopez ED, Fortney JJ. 2013. *Ap. J.* 776:2
- Lopez ED, Rice K. 2018. *MNRAS* 479:5303–11
- Lundkvist MS, Kjeldsen H, Albrecht S, et al. 2016. *Nat. Commun.* 7:11201
- Ma S, Mao S, Ida S, Zhu W, Lin DNC. 2016. *MNRAS* 461:L107–11
- Madhusudhan N. 2019. *Annu. Rev. Astron. Astrophys.* 57:617–63
- Malla SP, Stello D, Huber D, et al. 2020. *MNRAS* 496:5423–35
- Mao S. 2012. *Res. Astron. Astrophys.* 12:947–72
- Mao S, Paczynski B. 1991. *Ap. J. Lett.* 374:L37–40
- Masuda K, Winn JN, Kawahara H. 2020. *Astron. J.* 159:38
- Matsakos T, Königl A. 2016. *Ap. J. Lett.* 820:L8
- Mayor M, Marmier M, Lovis C, et al. 2011. arXiv:1109.2497
- Mayor M, Queloz D. 1995. *Nature* 378:355–59
- Mazeh T, Holczer T, Faigler S. 2016. *Astron. Astrophys.* 589:A75
- McArthur BE, Benedict GF, Barnes R, et al. 2010. *Ap. J.* 715:1203–20
- McDonald GD, Kreidberg L, Lopez E. 2019. *Ap. J.* 876:22
- Millholland S, Laughlin G. 2019. *Nat. Astron.* 3:424–33
- Millholland S, Wang S, Laughlin G. 2017. *Ap. J. Lett.* 849:L33
- Mills SM, Fabrycky DC. 2017. *Astron. J.* 153:45
- Mills SM, Howard AW, Petigura EA, et al. 2019. *Astron. J.* 157:198
- Mizuno H. 1980. *Prog. Theor. Phys.* 64:544–57
- Moe M, Kratter KM. 2019. *MNRAS*. Submitted. arXiv:1912.01699
- Moe M, Kratter KM, Badenes C. 2019. *Ap. J.* 875:61
- Moorhead AV, Ford EB, Morehead RC, et al. 2011. *Ap. J. Suppl.* 197:1
- Morbidelli A, Lambrechts M, Jacobson S, Bitsch B. 2015. *Icarus* 258:418–29
- Mordasini C, Alibert Y, Benz W. 2009. *Astron. Astrophys.* 501:1139–60
- Morton TD, Bryson ST, Coughlin JL, et al. 2016. *Ap. J.* 822:86
- Mróz P, Poleski R, Gould A, et al. 2020a. *Ap. J. Lett.* 903:L11
- Mróz P, Poleski R, Han C, et al. 2020b. *Astron. J.* 159:262
- Mróz P, Ryu YH, Skowron J, et al. 2018. *Astron. J.* 155:121
- Mróz P, Udalski A, Bennett DP, et al. 2019. *Astron. Astrophys.* 622:A201
- Mróz P, Udalski A, Skowron J, et al. 2017b. *Nature* 548:183–86
- Mulders GD, Mordasini C, Pascucci I, et al. 2019. *Ap. J.* 887:157
- Mulders GD, Pascucci I, Apai D. 2015a. *Ap. J.* 798:112
- Mulders GD, Pascucci I, Apai D. 2015b. *Ap. J.* 814:130
- Mulders GD, Pascucci I, Apai D. 2016. *Astron. J.* 152:187
- Mulders GD, Pascucci I, Apai D, Ciesla FJ. 2018. *Astron. J.* 156:24
- Munoz Romero CE, Kempton EMR. 2018. *Astron. J.* 155:134
- Murchikova L, Tremaine S. 2020. *Astron. J.* 160:160
- Nakagawa Y, Sekiya M, Hayashi C. 1986. *Icarus* 67:375–90
- Ngo H, Knutson HA, Hinkley S, et al. 2016. *Ap. J.* 827:8
- Nielsen EL, De Rosa RJ, Macintosh B, et al. 2019. *Astron. J.* 158:13
- Ofir A, Xie JW, Jiang CF, Sari R, Aharonson O. 2018. *Ap. J. Suppl.* 234:9
- Ormel CW. 2017. In *Formation, Evolution, and Dynamics of Young Solar Systems*, ed. M Pessah, O Gressel. *Ap. Space Sci. Libr.* 445:197–228. Cham, Switz.: Springer
- Ormel CW, Klahr HH. 2010. *Astron. Astrophys.* 520:A43



- Owen JE. 2019. *Annu. Rev. Earth Planet. Sci.* 47:67–90
- Owen JE, Lai D. 2018. *MNRAS* 479:5012–21
- Owen JE, Murray-Clay R. 2018. *MNRAS* 480:2206–16
- Owen JE, Wu Y. 2013. *Ap. J.* 775:105
- Owen JE, Wu Y. 2017. *Ap. J.* 847:29
- Pascucci I, Mulders GD, Gould A, Fernandes R. 2018. *Ap. J. Lett.* 856:L28
- Penny MT, Gaudi BS, Kerins E, et al. 2019. *Ap. J. Suppl.* 241:3
- Perryman M, Hartman J, Bakos GÁ, Lindegren L. 2014. *Ap. J.* 797:14
- Petigura EA, Howard AW, Marcy GW. 2013. *PNAS* 110:19273–78
- Petigura EA, Howard AW, Marcy GW, et al. 2017. *Astron. J.* 154:107
- Petigura EA, Marcy GW, Winn JN, et al. 2018. *Astron. J.* 155:89
- Petrovich C, Deibert E, Wu Y. 2019. *Astron. J.* 157:180
- Petrovich C, Tremaine S. 2016. *Ap. J.* 829:132
- Plavchan P, Barclay T, Gagné J, et al. 2020. *Nature* 582:497–500
- Plavchan P, Bilinski C, Currie T. 2014. *Publ. Astron. Soc. Pac.* 126:34–47
- Poleski R, Skowron J, Udalski A, et al. 2014. *Ap. J.* 795:42
- Pollack JB, Hubickyj O, Bodenheimer P, et al. 1996. *Icarus* 124:62–85
- Pu B, Lai D. 2019. *MNRAS* 488:3568–87
- Pu B, Lai D. 2020. arXiv:2008.05698
- Pu B, Wu Y. 2015. *Ap. J.* 807:44
- Quillen AC. 2011. *MNRAS* 418:1043–54
- Rasio FA, Ford EB. 1996. *Science* 274:954–56
- Raymond SN, Barnes R, Mandell AM. 2008. *MNRAS* 384:663–74
- Ricker GR, Winn JN, Vanderspek R, Latham DW, Bakos GÁ, et al. 2015. *J. Astron. Telesc. Instrum. Syst.* 1:014003
- Rogers JG, Owen JE. 2021. *MNRAS* 503:1526–42
- Ryu YH, Mróz P, Gould A, et al. 2021. *Astron. J.* 161:126
- Safronov VS. 1972. *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets*. Jerusalem: Israel Program Sci. Transl.
- Sanchis-Ojeda R, Rappaport S, Winn JN, et al. 2014. *Ap. J.* 787:47
- Sandford E, Kipping D, Collins M. 2019. *MNRAS* 489:3162–3173
- Santerne A, Moutou C, Tsantaki M, et al. 2016. *Astron. Astrophys.* 587:A64
- Santos NC, Israelian G, Mayor M. 2001. *Astron. Astrophys.* 373:1019–31
- Schlaufman KC, Lin DNC, Ida S. 2010. *Ap. J. Lett.* 724:L53–58
- Schlaufman KC, Winn JN. 2013. *Ap. J.* 772:143
- Schlaufman KC, Winn JN. 2016. *Ap. J.* 825:62
- Schlecker M, Mordasini C, Emsenhuber A, et al. 2020. *Astron. Astrophys.* In press. <https://doi.org/10.1051/0004-6361/202038554>
- Schlichting HE. 2014. *Ap. J. Lett.* 795:L15
- Seager S, Mallén-Ornelas G. 2003. *Ap. J.* 585:1038–55
- Shallue CJ, Vanderburg A. 2018. *Astron. J.* 155:94
- Shvartzvald Y, Maoz D, Udalski A, et al. 2016. *MNRAS* 457:4089–113
- Skidmore W, TMT Intl. Sci. Dev. Teams, Sci. Advis. Comm. 2015. *Res. Astron. Astrophys.* 15:1945–2140
- Sousa SG, Santos NC, Mayor M, et al. 2008. *Astron. Astrophys.* 487:373–81
- Steffen JH, Batalha NM, Borucki WJ, et al. 2010. *Ap. J.* 725:1226–41
- Steffen JH, Farr WM. 2013. *Ap. J. Lett.* 774:L12
- Steffen JH, Ragozzine D, Fabrycky DC, et al. 2012. *PNAS* 109:7982–87
- Stevenson DJ. 1982. *Planet. Space Sci.* 30:755–64
- Sumi T, Bennett DP, Bond IA, et al. 2010. *Ap. J.* 710:1641–53
- Sumi T, Kamiya K, Bennett DP, et al. 2011. *Nature* 473:349–52
- Suzuki D, Bennett DP, Ida S, et al. 2018. *Ap. J. Lett.* 869:L34
- Suzuki D, Bennett DP, Sumi T, et al. 2016. *Ap. J.* 833:145

- Szabó GM, Kiss LL. 2011. *Ap. J. Lett.* 727:L44
- Tabachnik S, Tremaine S. 2002. *MNRAS* 335:151–58
- Terquem C, Papaloizou JCB. 2019. *MNRAS* 482:530–49
- Teske J, Daz MR, Luque R, et al. 2020. *Astron. J.* 160:96
- Teske JK, Thorngren D, Fortney JJ, Hinkel N, Brewer JM. 2019. *Astron. J.* 158:239
- Thompson SE, Coughlin JL, Hoffman K, et al. 2018. *Ap. J. Suppl.* 235:38
- Tremaine S. 2015. *Ap. J.* 807:157
- Tremaine S, Dong S. 2012. *Astron. J.* 143:94
- Tu L, Johnstone CP, Güdel M, Lammer H. 2015. *Astron. Astrophys.* 577:L3
- Udalski A, Ryu YH, Sajadian S, et al. 2018. *Acta Astron.* 68:1–42
- Udalski A, Szymanski MK, Szymanski G. 2015a. *Acta Astron.* 65:1–38
- Udalski A, Yee JC, Gould A, et al. 2015b. *Ap. J.* 799:237
- Uehara S, Kawahara H, Masuda K, Yamada S, Aizawa M. 2016. *Ap. J.* 822:2
- Van Eylen V, Agentoft C, Lundkvist MS, et al. 2018. *MNRAS* 479:4786–95
- Van Eylen V, Albrecht S. 2015. *Ap. J.* 808:126
- Van Eylen V, Albrecht S, Huang X, et al. 2019. *Astron. J.* 157:61
- Venturini J, Alibert Y, Benz W. 2016. *Astron. Astrophys.* 596:A90
- Wang J, Fischer DA. 2015. *Astron. J.* 149:14
- Wang J, Fischer DA, Xie JW, Ciardi DR. 2015. *Ap. J.* 813:130
- Wang J, Xie JW, Barclay T, Fischer DA. 2014. *Ap. J.* 783:4
- Weidenschilling SJ. 1977. *Ap. Space Sci.* 51:153–58
- Weidenschilling SJ, Marzari F. 1996. *Nature* 384:619–21
- Weiss LM, Isaacson HT, Marcy GW, et al. 2018a. *Astron. J.* 156:254
- Weiss LM, Marcy GW, Petigura EA, et al. 2018b. *Astron. J.* 155:48
- Weiss LM, Petigura EA. 2020. *Ap. J. Lett.* 893:L1
- Winn JN. 2010. In *Exoplanets*, ed. Seager S. Tucson, AZ: Univ. Ariz. Press. arXiv:1001.2010
- Winn JN, Fabrycky DC. 2015. *Annu. Rev. Astron. Astrophys.* 53:409–47
- Winn JN, Sanchis-Ojeda R, Rappaport S. 2018. *New Astron. Rev.* 83:37–48
- Winn JN, Sanchis-Ojeda R, Rogers L, et al. 2017. *Astron. J.* 154:60
- Wisdom J. 1980. *Astron. J.* 85:1122–33
- Wittenmyer RA, Butler RP, Tinney CG, et al. 2016. *Ap. J.* 819:28
- Woolfson MM. 1993. *Q. J. R. Astron. Soc.* 34:1–20
- Wright JT, Marcy GW, Howard AW, et al. 2012. *Ap. J.* 753:160
- Wright JT, Upadhyay S, Marcy GW, et al. 2009. *Ap. J.* 693:1084–99
- Wright JT, Veras D, Ford EB, et al. 2011. *Ap. J.* 730:93
- Wu Y. 2019. *Ap. J.* 874:91
- Wu Y, Lithwick Y. 2013. *Ap. J.* 772:74
- Xie JW, Dong S, Zhu Z, Huber D, Zheng Z, et al. 2016. *PNAS* 113:11431–35
- Xie JW, Wu Y, Lithwick Y. 2014. *Ap. J.* 789:165
- Xuan JW, Wyatt MC. 2020. *MNRAS* 497:2096–118
- Yalinewich A, Petrovich C. 2020. *Ap. J. Lett.* 892:L11
- Yang JY, Xie JW, Zhou JL. 2020. *Astron. J.* 159:164
- Yee JC, Gould A, Beichman C, et al. 2015. *Ap. J.* 810:155
- Yee JC, Zang W, Udalski A. 2021. arXiv:2101.04696
- Youdin AN. 2011. *Ap. J.* 742:38
- Youdin AN, Goodman J. 2005. *Ap. J.* 620:459–69
- Zang W, Han C, Kondo I, et al. 2021. *Res. Astron. Astrophys.* Submitted. arXiv:2103.01896
- Zapatero Osorio MR, Béjar VJS, Martn EL, et al. 2000. *Science* 290:103–7
- Zhao G, Zhao Y-H, Chu Y-Q, Jing Y-P, Deng L-C. 2012. *Res. Astron. Astrophys.* 12:723–34
- Zhou JL, Lin DNC, Sun YS. 2007. *Ap. J.* 666:423–35
- Zhu W. 2019. *Ap. J.* 873:8
- Zhu W. 2020. *Astron. J.* 159:188

- Zhu W, Dai F, Masuda K. 2018a. *Res. Notes Am. Astron. Soc.* 2:160
- Zhu W, Petrovich C, Wu Y, Dong S, Xie J. 2018b. *Ap. J.* 860:101
- Zhu W, Wang J, Huang C. 2016. *Ap. J.* 832:196
- Zhu W, Wu Y. 2018. *Astron. J.* 156:92
- Ziegler C, Law NM, Baranec C, et al. 2018. *Astron. J.* 155:161
- Zink JK, Christiansen JL, Hansen BMS. 2019. *MNRAS* 483:4479–94
- Zinzi A, Turrini D. 2017. *Astron. Astrophys.* 605:L4
- Zong W, Fu JN, De Cat P, et al. 2018. *Ap. J. Suppl.* 238:30



# Contents

The Journey of a Radio Astronomer: Growth of Radio Astronomy in India <i>Govind Swarup</i> .....	1
Tidal Disruption Events <i>Suvi Gezari</i> .....	21
Microarcsecond Astrometry: Science Highlights from <i>Gaia</i> <i>Anthony G.A. Brown</i> .....	59
Observational Constraints on Black Hole Spin <i>Christopher S. Reynolds</i> .....	117
First Multimessenger Observations of a Neutron Star Merger <i>Raffaella Margutti and Ryan Chornock</i> .....	155
Transneptunian Space <i>Brett Gladman and Kathryn Volk</i> .....	203
Wave Dark Matter <i>Lam Hui</i> .....	247
Exoplanet Statistics and Theoretical Implications <i>Wei Zbu and Subo Dong</i> .....	291
Evolution and Mass Loss of Cool Aging Stars: A Daedalean Story <i>Leen Decin</i> .....	337
New Insights into Classical Novae <i>Laura Chomiuk, Brian D. Metzger, and Ken J. Shen</i> .....	391
Carrington Events <i>Hugh S. Hudson</i> .....	445

## Indexes

Cumulative Index of Contributing Authors, Volumes 48–59 .....	479
Cumulative Index of Article Titles, Volumes 48–59 .....	482

## Errata

An online log of corrections to *Annual Review of Astronomy and Astrophysics* articles may be found at <http://www.annualreviews.org/errata/astro>