MIT Open Access Articles

TESTS OF IN SITU FORMATION SCENARIOS FOR COMPACT MULTIPLANET SYSTEMS

Massachusetts Institute of Technology

TESTS OF IN SITU FORMATION SCENARIOS FOR COMPACT MULTIPLANET SYSTEMS

Kevin C. Schlaufman[1]
Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; kschlauf@mit.edu

ABSTRACT

Kepler has identified over 600 multiplanet systems, many of which have several planets with orbital distances smaller than that of Mercury. Because these systems may be difficult to explain in the paradigm of core accretion and disk migration, it has been suggested that they formed in situ within protoplanetary disks with high solid surface densities. The strong connection between giant planet occurrence and stellar metallicity is thought to be linked to enhanced solid surface densities in disks around metal-rich stars, so the presence of a giant planet can be a sign of planet formation in a high solid surface density disk. I formulate quantitative predictions for the frequency of long-period giant planets in these in situ models by translating the proposed increase in disk mass into an equivalent metallicity enhancement. I rederive the scaling of giant planet occurrence with metallicity as $P_{\mathrm{gp}} = 0.05^{+0.02}_{-0.02} \times 10^{(2.1\pm0.4)\mathrm{[M/H]}} = 0.08^{+0.02}_{-0.03} \times 10^{(2.3\pm0.4)\mathrm{[Fe/H]}}$ and show that there is significant tension between the frequency of giant planets suggested by the minimum mass extrasolar nebula scenario and the observational upper limits. Consequently, high-mass disks alone cannot explain the observed properties of the close-in *Kepler* multiplanet systems and therefore migration is still important. More speculatively, I combine the metallicity scaling of giant planet occurrence with small planet occurrence rates to estimate the number of solar system analogs in the Galaxy. I find that in the Milky Way there are perhaps 4×10^6 true solar system analogs with an FGK star hosting both a terrestrial planet in the habitable zone and a long-period giant planet companion.

Key words: Galaxy: general – methods: statistical – planetary systems – planets and satellites: detection – planets and satellites: formation – stars: statistics

Online-only material: color figures

1. INTRODUCTION

Kepler has discovered many multiplanet systems with several planets with orbital periods $P < 50$ days.[2] Indeed, 40% of solar-type stars in the *Kepler* field have at least one planet with $P < 50$ days (e.g., Fressin et al. 2013). Even though these systems differ from the solar system, their apparent ubiquity suggests that they may represent a frequent outcome of planet formation.

In the traditional minimum-mass solar nebula (MMSN) scenario, there is probably insufficient solid material in protoplanetary disks to form the *Kepler* multiplanet systems where they are observed today (Weidenschilling 1977; Hayashi 1981). Instead, formation further out in the parent protoplanetary disk combined with subsequent inward migration has been suggested as one possible formation channel for this class of system (e.g., Alibert et al. 2006). The apparent excess of planets just outside of mean-motion resonances may also support formation in the inward migration scenario (e.g., Lissauer et al. 2011; Fabrycky et al. 2012). However, the rate and even direction of migration is thought to sensitively depend on the unknown thermodynamic state of the disk (e.g., Paardekooper et al. 2010; Kley & Nelson 2012).

Alternative models of in situ formation in disks with solid surface densities enhanced beyond the MMSN expectation have also been suggested to explain the ubiquity of close-in multiple systems. In the minimum-mass extrasolar nebula (MMEN) scenario of Chiang & Laughlin (2013), the *Kepler* multiplanet systems formed in protoplanetary disks that were about six times

more massive than those envisioned in the MMSN scenario. In that picture, the more massive MMEN disks describe the typical protoplanetary disk in the Galaxy, while the less massive MMSN disk is the outlier. I illustrate the MMEN in Figure 1. On the other hand, Hansen & Murray (2012) invoke the rapid inward migration of planetesimals into the inner regions of the disk. The enhanced solid surface density of the inner disk then naturally leads to the in situ formation of planetary systems closely resembling those observed by *Kepler* (Hansen & Murray 2013).

Both models of in situ planet formation described above provide useful, fresh looks at planet formation. As I will show, both models are also amenable to quantitative tests. At face value, both models make qualitative predictions for the formation of long-period giant planets. All else being equal, a disk with higher solid surface density in the giant planet forming region has a better chance of forming a giant planet than does a disk with lower solid surface density in the giant planet forming region (e.g., Hubickyj et al. 2005). Consequently, Chiang & Laughlin (2013) extrapolated their model to larger semimajor axes and suggested that the enhanced solid surface density in the MMEN scenario should lead to the efficient formation of giant planets outside of 1 AU. In contrast, the concentration of a significant amount of a disk's solid material in the inner disk as suggested by Hansen & Murray (2012) should lead to inefficient formation of giant planets outside of 1 AU. Unfortunately, it is not currently possible to directly assess with either the transit or radial velocity (RV) technique the frequency of long-period giant planets in the observed *Kepler* multiple systems themselves.

However, it is possible to indirectly infer the frequency of long-period giant planets in at least two ways. First, the frequency can be characterized by proxy, a technique in which

---

[1] Kavli Fellow.
[2] See, for example, Borucki et al. (2011a, 2011b), Batalha et al. (2013), and Burke et al. (2014).
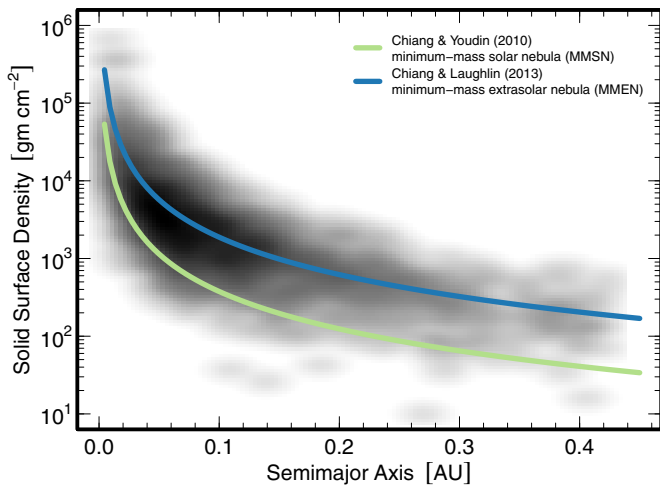
Figure 1. Solid surface density profile Σ_{solid} of the Chiang & Laughlin (2013) minimum-mass extrasolar nebula (MMEN). The background shading indicates the density of points in the semimajor axis–solid surface density diagram for *Kepler* objects of interest (KOIs). I define the solid surface density of a KOI as $\Sigma_{\mathrm{solid}} \equiv M_p/(2\pi a_p^2)$. I compute the mass of each KOI by assuming $M_p = R_p^{2.06}$ (e.g., Lissauer et al. 2011) and each semimajor axis a_p using the observed period P and assuming a $1\,M_\odot$ host star. The green curve shows the standard Hayashi (1981) minimum-mass solar nebula (MMSN) as parameterized by Chiang & Youdin (2010), while the blue curve shoes the fiducial MMEN of Chiang & Laughlin (2013). There is insufficient solid material in the MMSN nebula to form the observed KOIs in situ, so migration of solids is required to explain the KOIs. On the other hand, there is sufficient solid material present in the MMEN scenario to form the KOIs in situ.

(A color version of this figure is available in the online journal.)

objects in the solar neighborhood that can be studied in detail stand in for the more distant *Kepler* objects. In this case, the close-in multiple systems of low-mass planets discovered in the solar neighborhood with the RV technique are a proxy for the more distant *Kepler* multiple systems. One can use the published completeness limits of the RV surveys to establish upper limits on the frequency of long-period giant planets in those systems, then compare that upper limit to quantitative predictions of the in situ models of planet formation. Second, it well known that giant planet host stars are preferentially metal enriched (e.g., Santos et al. 2004; Fischer & Valenti 2005). Therefore, the absence of a metallicity effect in solar-type hosts of *Kepler* multiple systems can be used to determine an upper limit on the frequency of giant planets in these systems.

In this paper, I compare statistical upper limits on the frequency of long-period giant planets in the *Kepler* multiple systems with quantitative predictions of the in situ models of planet formation. I find that there is significant tension between the derived upper limits and the expectation from the MMEN scenario, though current samples are not large enough to constrain the Hansen & Murray (2012) scenario. I describe my sample selection in Section 2, I detail my statistical analyses in Section 3, I discuss the results and implications in Section 4, and I summarize my findings in Section 5.

2. SAMPLE DEFINITION

I select *Kepler* multiple planet systems having at least two planets with $R_P < 5\,R_\oplus$ from the *Kepler* CasJobs database[3]

---

[3] http://mastweb.stsci.edu/kplrcasjobs/guide.aspx

using the query given in the Appendix. I focus exclusively on exoplanet systems orbiting solar-type stars, because the MMEN scenario is scaled from the solar nebula and because there is both a theoretical expectation and observational evidence that the planet formation process changes for low-mass stars (e.g., Laughlin et al. 2004; Ida & Lin 2005; Endl et al. 2006; Butler et al. 2006; Bonfils et al. 2013). For that reason, I select systems orbiting solar-type stars with $0.22 < J - H < 0.62$, $0.00 < H - K < 0.10$, and $0.22 < J - K < 0.72$ roughly corresponding to spectral types in the range F5–K5 (Covey et al. 2007). From here, I refer to the systems in this sample as the *Kepler* multiplanet systems.

Likewise, I select systems of exoplanets discovered with the RV technique that have at least two planets with $M \sin i \leqslant 0.1\,M_{\mathrm{Jup}} = 31.8\,M_\oplus$ from both http://exoplanets.eu and http://exoplanets.org (Schneider et al. 2011; Wright et al. 2011). For each planet host star, I obtain *Hipparcos* parallaxes and $B - V$ colors from van Leeuwen (2007) and apparent *Tycho-2* V-band magnitudes from Høg et al. (2000). I transform *Tycho-2* B_T and V_T magnitudes into approximate Johnson–Cousins V-band magnitudes using the relation $V = V_T - 0.090(B_T - V_T)$. I then select exoplanet systems orbiting solar-type stars with $0.44 < B - V < 1.15$ and $3.5 < M_V < 7.4$, roughly corresponding to spectral types in the range F5–K5 (Binney & Merrifield 1998). I give the exoplanet systems and host star properties that result from this selection in Tables 1 and 2. From here, I refer to the systems in this sample as the RV multiplanet systems. I plot both samples in Figure 2. I make no distinction between mini-Neptunes and super-Earths despite the possibility that their origins may be unique.

It is important to establish the utility of the RV multiplanet systems as a proxy for the *Kepler* multiplanet systems. Radial velocity surveys measure the stellar reflex velocity, which for a given host star mass is a function of planet mass, orbital period, eccentricity, and inclination. On the other hand, transit surveys measure transit depth, which for a given host star radius is a function of planet radius. Transit probability also biases transit surveys toward the shortest-period planets in an underlying population. Previous analyses have explored this relationship and discovered that while the two populations are broadly consistent, there may be differences (e.g., Wolfgang & Laughlin 2012). Connecting the two populations in this way requires an accurate and precise mass–radius relationship valid at all periods as well as knowledge of the orbital inclinations of the RV multiplanet systems. Despite the progress made for short-period planets, a general empirical mass–radius relation is not yet available (e.g., Marcy et al. 2014; Weiss & Marcy 2014). The orbital inclinations of the RV multiplanet systems are also unknown.

However, the similarity of the *Kepler* and RV multiplanet systems can be established in another way. If the same planet formation process is responsible for the formation of both the *Kepler* and the RV multiplanet systems, then the orbital spacing of the two populations should be similar. In a planetary system, the ratio of the separation between two planets to the sum of the Hill radii of two planets is

$$S_{\mathrm{Hill}} = \frac{a_1 - a_2}{R_{H,1} + R_{H,2}}, \qquad (1)$$

where $a_1 > a_2$. If $P_i^2 = [(4\pi^2)/(GM_{\mathrm{tot}})]a_i^3$ and

Table 1
Multiple Low-mass Planet Systems

Name	System	P (days)	e	K (m s^{-1})	$m \sin i$ ($M_\oplus$)	Reference
HD 1461 b	HD 1461	5.773	0.14	2.44	7.6	Rivera et al. (2010)
HD 1461 c	HD 1461	13.5	0	1.57	5.9	Mayor et al. (2011)
HD 10180 c	HD 10180	5.76	0.08	4.54	13	Lovis et al. (2011)
HD 10180 d	HD 10180	16.36	0.14	2.93	12	Lovis et al. (2011)
HD 10180 e	HD 10180	49.75	0.06	4.25	25	Lovis et al. (2011)
HD 10180 f	HD 10180	122.7	0.13	2.95	24	Lovis et al. (2011)
HD 10180 g	HD 10180	602	0	1.56	21	Lovis et al. (2011)
HD 10180 h	HD 10180	2248	0.15	3.11	66	Lovis et al. (2011)
HD 13808 b	HD 13808	14.18	0.17	3.53	10	Mayor et al. (2011)
HD 13808 c	HD 13808	53.83	0.43	2.81	11	Mayor et al. (2011)
HD 20003 b	HD 20003	11.85	0.4	4.03	12	Mayor et al. (2011)
HD 20003 c	HD 20003	33.82	0.16	2.95	13	Mayor et al. (2011)
HD 20794 b	HD 20794	18.32	0	0.83	2.7	Pepe et al. (2011)
HD 20794 c	HD 20794	40.11	0	0.56	2.4	Pepe et al. (2011)
HD 20794 d	HD 20794	90.31	0	0.85	4.7	Pepe et al. (2011)
HD 20781 b	HD 20781	29.15	0.11	3.03	12	Mayor et al. (2011)
HD 20781 c	HD 20781	85.13	0.28	2.88	16	Mayor et al. (2011)
HD 21693 b	HD 21693	22.66	0.26	2.73	10	Mayor et al. (2011)
HD 21693 c	HD 21693	53.88	0.24	4.02	21	Mayor et al. (2011)
HD 31527 b	HD 31527	16.55	0.13	3.01	12	Mayor et al. (2011)
HD 31527 c	HD 31527	51.28	0.11	2.83	16	Mayor et al. (2011)
HD 31527 d	HD 31527	274.5	0.38	1.79	17	Mayor et al. (2011)
HD 39194 b	HD 39194	5.636	0.2	1.95	3.7	Mayor et al. (2011)
HD 39194 c	HD 39194	14.02	0.11	2.26	5.9	Mayor et al. (2011)
HD 39194 d	HD 39194	33.94	0.2	1.49	5.2	Mayor et al. (2011)
HD 40307 b	HD 40307	4.312	0	1.97	4.1	Mayor et al. (2009)
HD 40307 c	HD 40307	9.62	0	2.47	6.7	Mayor et al. (2009)
HD 40307 d	HD 40307	20.46	0	2.55	8.9	Mayor et al. (2009)
HD 51608 b	HD 51608	14.07	0.15	4.10	13	Mayor et al. (2011)
HD 51608 c	HD 51608	95.42	0.41	3.25	18	Mayor et al. (2011)
HD 69830 b	HD 69830	8.667	0.1	3.51	10	Lovis et al. (2006)
HD 69830 c	HD 69830	31.56	0.13	2.66	12	Lovis et al. (2006)
HD 69830 d	HD 69830	197	0.07	2.2	18	Lovis et al. (2006)
HD 93385 b	HD 93385	13.19	0.15	2.21	8.4	Mayor et al. (2011)
HD 93385 c	HD 93385	46.02	0.24	1.82	10	Mayor et al. (2011)
HD 96700 b	HD 96700	8.126	0.1	3.02	9	Mayor et al. (2011)
HD 96700 c	HD 96700	103.5	0.37	1.98	13	Mayor et al. (2011)
HD 109271 b	HD 109271	7.854	0.25	5.6	17	Lo Curto et al. (2013)
HD 109271 c	HD 109271	30.93	0.15	4.9	24	Lo Curto et al. (2013)
61 Vir b	61 Vir	4.215	0.12	2.12	5.1	Vogt et al. (2010)
61 Vir c	61 Vir	38.02	0.14	2.12	11	Vogt et al. (2010)
61 Vir d	61 Vir	123	0.35	3.25	23	Vogt et al. (2010)
HD 134606 b	HD 134606	12.08	0.15	2.68	9.3	Mayor et al. (2011)
HD 134606 c	HD 134606	59.52	0.29	2.17	12	Mayor et al. (2011)
HD 134606 d	HD 134606	459.3	0.46	3.66	38	Mayor et al. (2011)
HD 136352 b	HD 136352	11.58	0.18	1.77	5.3	Mayor et al. (2011)
HD 136352 c	HD 136352	27.58	0.16	2.82	11	Mayor et al. (2011)
HD 136352 d	HD 136352	106.7	0.43	1.68	9.5	Mayor et al. (2011)
HD 192310 b	HD 192310	74.72	0.13	3	17	Pepe et al. (2011)
HD 192310 c	HD 192310	525.8	0.32	2.27	23	Pepe et al. (2011)
HD 215152 b	HD 215152	7.282	0.34	1.26	2.8	Mayor et al. (2011)
HD 215152 c	HD 215152	10.87	0.38	1.26	3.1	Mayor et al. (2011)

$$R_{H,i} = \left(\frac{M_i}{3M_*}\right)^{1/3} a_i, \qquad (2)$$

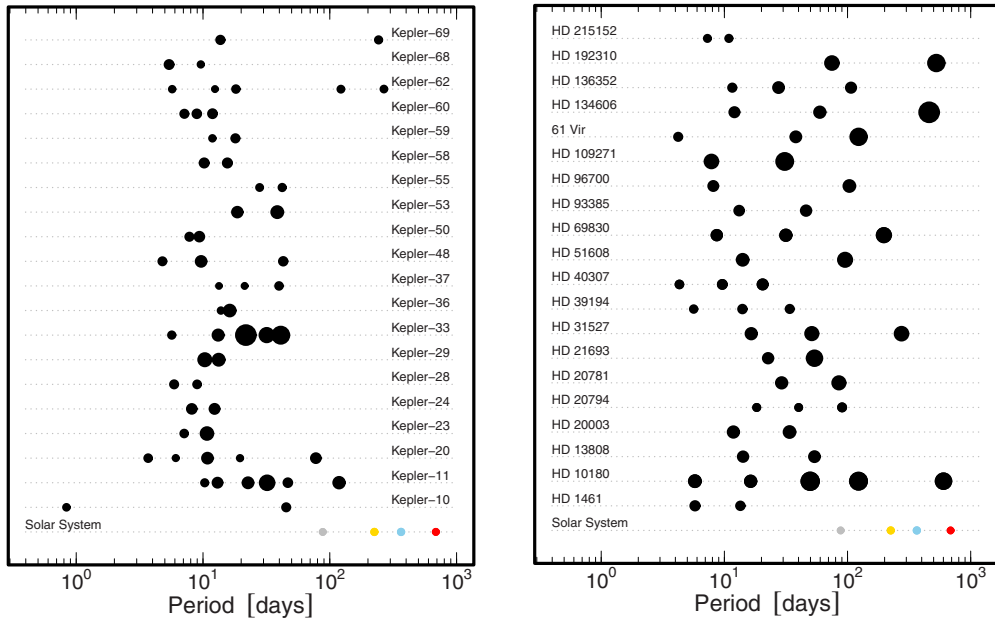$$= \left(\frac{\rho_i}{3\rho_*}\right)^{1/3} \left(\frac{r_i}{r_*}\right) a_i, \qquad (3)$$

$$= \left(\frac{GM_{\text{tot}}}{4\pi^2}\right)^{1/3} \left(\frac{\rho_i}{3\rho_*}\right)^{1/3} \left(\frac{r_i}{r_*}\right) P_i^{2/3}, \qquad (4)$$

then

$$S_{\text{Hill}} = \frac{a_1 - a_2}{R_{H,1} + R_{H,2}}, \qquad (5)$$

$$= \frac{P_1^{2/3} - P_2^{2/3}}{\left[P_1^{2/3}(\rho_1/3\rho_*)^{1/3}(r_1/r_*) + P_2^{2/3}(\rho_2/3\rho_*)^{1/3}(r_2/r_*)\right]}. \qquad (6)$$

Both P_i and r_i/r_* are directly observable, so using Equation (6) for the *Kepler* multiplanet systems only requires assuming

Figure 2. Multiplanet system architectures. Left: validated *Kepler* multiplanet systems. The size of each planet is proportional to its estimated radius. The terrestrial planets in the solar system are included for scale. Right: systems of multiple Neptune-mass planets discovered with the radial velocity technique (RV multiplanet systems). The size of each planet is proportional to its estimated minimum mass. The terrestrial planets in the solar system are included for scale. The architectures of the *Kepler* and RV multiplanet systems are similar, indicating that the properties of RV multiplanet systems can reasonably be used as proxies for the properties of the *Kepler* multiplanet systems.

(A color version of this figure is available in the online journal.)

Table 2
Host Stars of Multiple Low-mass Planet Systems

Name	HIP	HD	M_V (mag)	$B - V$ (mag)	N_p
HD 1461	1499	1461	4.63	0.67	2
HD 10180	7599	10180	4.37	0.63	6
HD 13808	10301	13808	6.10	0.87	2
HD 20003	14530	20003	5.17	0.77	2
HD 20794	15510	20794	5.36	0.71	3
HD 20781	15526	20781	5.71	0.82	2
HD 21693	16085	21693	5.40	0.76	2
HD 31527	22905	31527	4.56	0.61	3
HD 39194	27080	39194	6.02	0.76	3
HD 40307	27887	40307	6.59	0.94	3
HD 51608	33229	51608	5.46	0.77	2
HD 69830	40693	69830	5.47	0.75	3
HD 93385	52676	93385	4.37	0.60	2
HD 96700	54400	96700	4.47	0.61	2
HD 109271	61300	109271	4.11	0.66	2
61 Vir	64924	115617	5.08	0.71	3
HD 134606	74653	134606	4.75	0.74	3
HD 136352	75181	136352	4.80	0.64	3
HD 192310	99825	192310	5.98	0.88	2
HD 215152	112190	215152	6.47	0.97	2

an average density. For the RV multiplanet systems, I use Equations (1) and (2) and the published values of a, $m \sin i$, and M_* for each system.

I plot the Hill-spacing distributions for both the *Kepler* and the RV multiplanet systems in Figure 3, separating two- and three-planet systems and assuming different average densities. I compare the distributions using the Anderson–Darling test (e.g., Stephens 1974; Conover 1999; Hou et al. 2009). If the typical exoplanet density in the *Kepler* sample resembles Neptune, then the Hill-spacing distributions of the *Kepler* and RV multiplanet systems are indistinguishable. Alternatively,

if the typical density in the *Kepler* sample resembles the Earth, then it is possible that the three-planet systems in the *Kepler* multiplanet sample are more compact than the three-planet systems in the RV multiplanet sample. Under either density assumption, the two-planet systems are statistically indistinguishable. Fortunately for this analysis, most of the planets in the *Kepler* and RV multiplanet systems have $P >$ 10 days. In this range of orbital period, most planets in the *Kepler* sample are thought to have a Neptune-like average density $\rho \approx 1$ g cm^{-3} (e.g., Wu & Lithwick 2013). Therefore, the statistical similarity of the Hill-spacing distributions of the two samples supports the utility of the RV multiplanet systems as a proxy for the *Kepler* multiplanet systems.

3. ANALYSIS

3.1. Giant Planet Formation

It is well established that giant planets occur more frequently around metal-rich stars (e.g., Santos et al. 2004; Fischer & Valenti 2005). The high metallicity of a star is interpreted as evidence that its parent disk was enriched in dust. A dust-enriched protoplanetary disk is thought to be more likely to form the $\approx 10\,M_\oplus$ core necessary for giant planet formation in the few million years available before the disk disappears. Giant planets are also less frequently found around M dwarfs than solar-type stars (e.g., Endl et al. 2006; Butler et al. 2006; Bonfils et al. 2013). M dwarfs presumably formed from lower-mass disks than solar-type stars, and the reduced amount of dust present in a low-mass disk is thought to make it more difficult to assemble a core. Recent observational evidence that disk mass scales approximately linearly with host star mass below 1 $M_\odot$ supports this view (Andrews et al. 2013).

In short, giant planet formation efficiency is proportional to the solid surface density in a protoplanetary disk $\Sigma_{\rm solid}$, where

$$\Sigma_{\rm solid} \propto f_{\rm dust} M_{\rm disk} \propto Z M_{\rm disk}. \qquad (7)$$
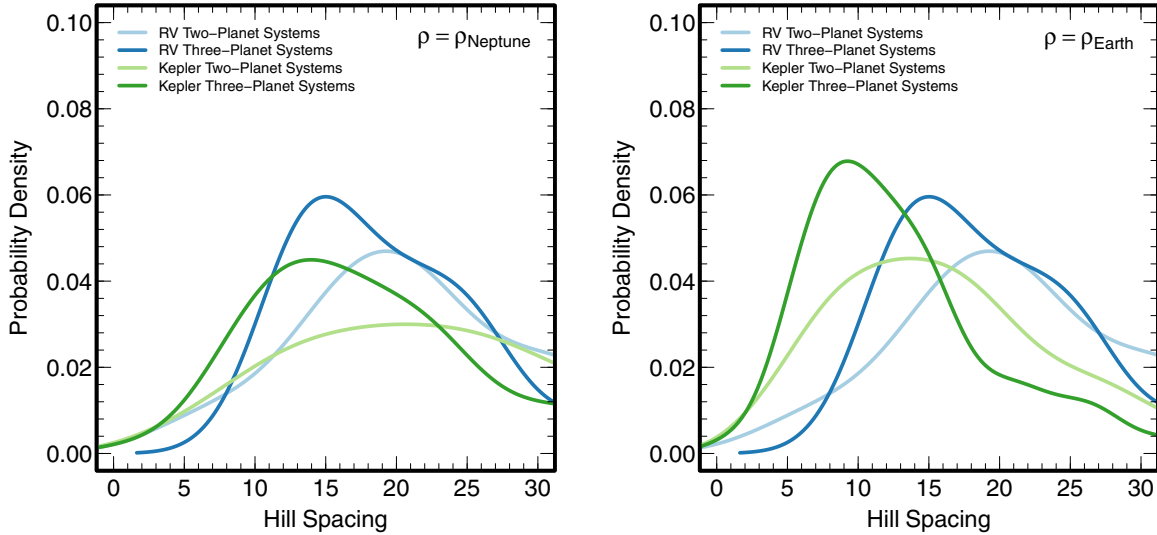
Figure 3. Hill-spacing distributions of *Kepler* and RV multiplanet systems. Left: Hill-spacing distribution assuming that the typical density ρ of planet candidates in the *Kepler* multiplanet systems is $\rho = \rho_{\mathrm{Neptune}} = 1.6$ g cm^{-3}. Right: Hill-spacing distribution assuming $\rho = \rho_{\oplus} = 5.5$ g cm^{-3}. If $\rho = \rho_{\mathrm{Neptune}}$, then both distributions are statistically indistinguishable, as the *p*-value from the Anderson–Darling test comparing the *Kepler* and RV multiplanet systems is 0.4 for both the two-planet and three-planet systems. If $\rho = \rho_{\oplus}$, then the *p*-value from the Anderson–Darling test comparing the *Kepler* and RV multiplanet systems is 0.1 for the two-planet systems and 0.002 for the three-planet systems.

(A color version of this figure is available in the online journal.)

The first proportionality is true by construction. The second is true because observations of the LMC and Milky Way show that the dust-to-gas fraction f_{dust} scales approximately linearly with metal mass fraction Z (Gordon et al. 2003). If the MMEN disk can be extrapolated beyond 1 AU, then the six times increase in disk mass in the MMEN scenario relative to the MMSN scenario should have an equivalent effect on giant planet formation efficiency as increasing Z by a factor of six, or increasing [M/H] by 0.78 dex.

The pioneering work of Fischer & Valenti (2005) showed that the probability of giant planet occurrence P_{gp} is

$$P_{\mathrm{gp}} = 0.03 \times 10^{2.0[\mathrm{Fe/H}]}. \qquad (8)$$

They argued that the dependence of P_{gp} on the square of the number of iron atoms present was the natural expectation from the collisional agglomeration of dust grains into planetesimals. However, their result lacked an uncertainty estimate and may have been affected by the need to bin their data. Moreover, many new, longer-period giant planets have been discovered in the interim, so a new calculation is timely. For those reasons, I use logistic regression to rederive the scaling of giant planet occurrence with metallicity, as logistic regression both avoids the need to bin the data and naturally produces an error estimate on the scaling.[4]

As input, I use a sample of 1111 FGK field and planet host stars from the HARPS GTO planet search program from Adibekyan et al. (2012b).[5] Stellar parameters and abundances for each star in the catalog have been homogeneously derived using the techniques presented in Sousa et al. (2008, 2011b, 2011a). I cross-match all of the stars with the *Hipparcos*, *Tycho-2*, and exoplanets.org catalogs using TOPCAT[6] (Taylor 2005). I retain only those stars with distance $d < 50$ pc from

the Sun, parallaxes precise to 20%, $0.44 < B - V < 1.15$, and $3.5 < M_V < 7.4$. Finally, I code as giant planet host stars those stars with planets with RV semiamplitude $K > 20$ m s^{-1} and $P < 4$ yr, because the completeness of the survey is very high in that range of parameter space. The end result is a sample of 620 solar-type stars, 44 of which host at least one giant planet.

The probability of giant planet occurrence must be in the interval $0 \leqslant P_{\mathrm{gp}} \leqslant 1$. Linear regression is unsuitable for the prediction of probabilities, because the linear regression equation $y = \beta_0 + \sum \beta_i x_i$ is not bounded between 0 and 1. Instead, logistic regression makes use of the logistic function to predict the probability

$$P(Y) = \frac{e^{\beta_0 + \sum \beta_i x_i}}{1 + e^{\beta_0 + \sum \beta_i x_i}}. \qquad (9)$$

I plot the logistic function in Figure 4. It takes any real value x and maps it into the interval $0 \leqslant y \leqslant 1$, satisfying the requirement that the probability of an event be between 0 and 1.

The logistic function is nonlinear. As a result, logistic regression works with the natural logarithm of the odds ratio

$$\log\left[\frac{P(Y)}{1 - P(Y)}\right] = \beta_0 + \sum \beta_i x_i, \qquad (10)$$

or the logit function, which is linear in the coefficients β_i. The coefficients β_i can then be fit numerically. For $i > 0$, the interpretation of the coefficient β_i is that a one unit change in x_i changes the log odds ratio of the probability of the event by a factor of β_i, or the probability of the event itself by e^{β_i}. The coefficient β_0 is the log odds of the event when all predictors $x_i = 0$.

In this context, I calculate the response of P_{gp} to only one predictor: either $x_1 = [\mathrm{M/H}]$ or $x_1 = [\mathrm{Fe/H}]$. I use the `glm` function in R[7] to compute a logistic regression model (R Core Team 2013). I build models predicting the effect of both

---

[4] For a pedagogical introduction to logistic regression, see Chapter 12 of Chatterjee et al. (2000).
[5] I arrive at quantitatively similar results if I instead use the Valenti & Fischer (2005) SPOCS sample.
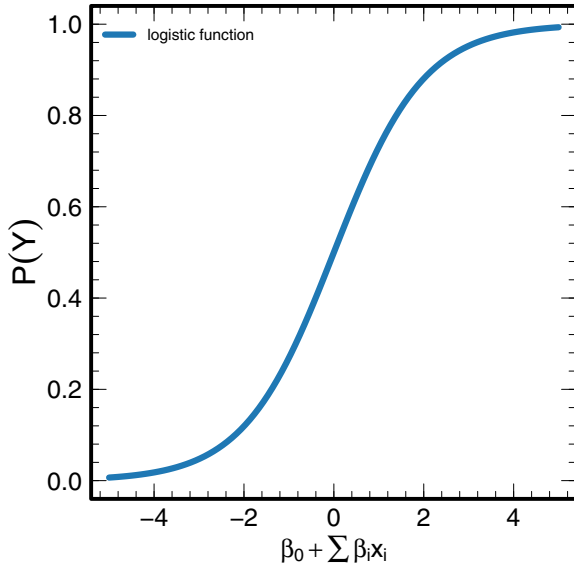[6] http://www.star.bristol.ac.uk/~mbt/topcat/

[7] http://www.R-project.org/

Figure 4. Logistic function.
(A color version of this figure is available in the online journal.)

[M/H] and [Fe/H] on the probability of giant planet occurrence. I compute [M/H] assuming the solar abundances from Asplund et al. (2005). I find that

$$P_{gp}([M/H]) \propto 10^{(2.1\pm0.4)\Delta[M/H]}, \qquad (11)$$

$$P_{gp}([Fe/H]) \propto 10^{(2.3\pm0.4)\Delta[Fe/H]}. \qquad (12)$$

To determine the absolute probability of giant planet occurrence at $[M/H] = [Fe/H] = 0$, I calculate the fraction of stars with giant planets in the sample in the ranges $-0.05 < [M/H] < 0.05$ and $-0.05 < [Fe/H] < 0.05$. I use bootstrap resampling to determine confidence intervals, and I find that

near $[M/H] = [Fe/H] = 0$

$$P_{gp}([M/H]) = 0.05^{+0.02}_{-0.02} \times 10^{(2.1\pm0.4)[M/H]}, \qquad (13)$$

$$P_{gp}([Fe/H]) = 0.08^{+0.02}_{-0.03} \times 10^{(2.3\pm0.4)[Fe/H]}. \qquad (14)$$

I plot the result in Figure 5.

These results suggest that if the MMEN disk can be extrapolated past 1 AU, then the probability of giant planet occurrence in the *Kepler* multiplanet systems should be $P_{gp}(0.78) \approx 1$. Given that these giant planets would be found at $a \gtrsim 1$ AU, they would only rarely be observed to transit and therefore be nearly invisible to the transit technique. However, these planets would likely be detectable with the RV technique.

3.2. Stability and Completeness Constraints

The analysis in Section 3.1 indicates that practically every *Kepler* multiplanet system formed in the extrapolated MMEN scenario should have a long-period giant planet. However, it is currently impractical to search for long-period giant planets in the *Kepler* multiplanet systems themselves. Instead, I use the RV multiplanet systems in Table 2 as a proxy and use the published RV completeness estimates for those systems to derive a constraint on the frequency of long-period giant planets in the *Kepler* multiplanet systems.

The existence and stability of the *Kepler* multiplanet systems with orbital periods $P < 50$ days precludes the existence of giant planets with similar orbital periods. On the other hand, longer-period giant planets are not prohibited by stability arguments. A multiplanet system is likely to be stable if its planets are separated by 10 or more mutual Hill radii $R_H = [M_p/(3M_*)]^{1/3} a$ (e.g., Chambers et al. 1996; Smith & Lissauer 2009). For a system of multiple planets orbiting a $1 M_\odot$ star with its outermost Neptune-mass planet $M_1 = 17.147 M_\oplus$ at an
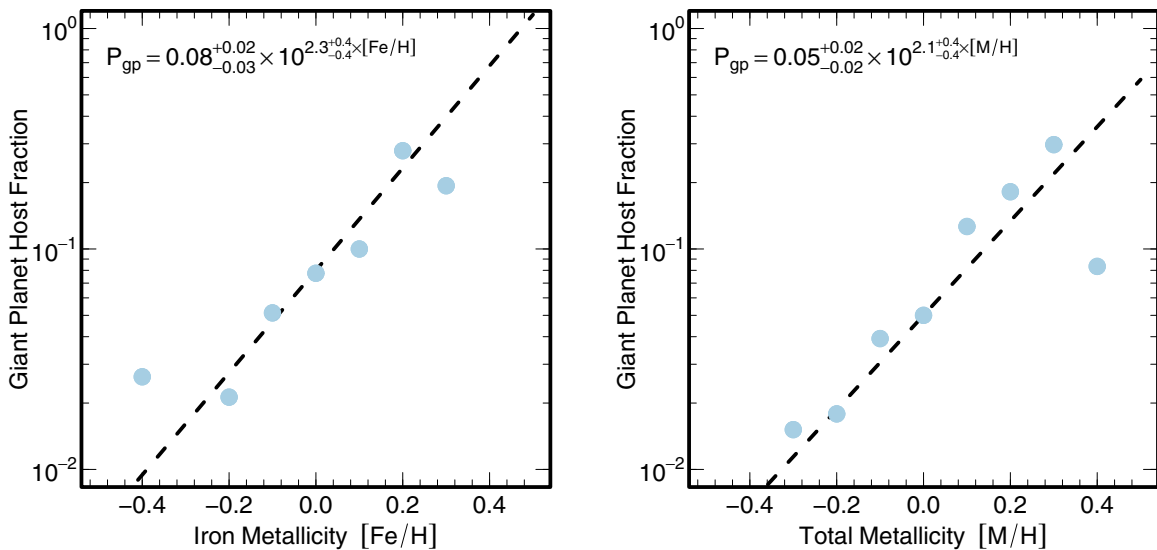

Figure 5. Effect of host star composition on giant planet occurrence. I use logistic regression to estimate the effect of composition on giant planet occurrence, so the arbitrary binning reflected in the plots does not affect the derived scaling. Left: the effect of iron metallicity on giant planet occurrence. There is a hint that giant planet occurrence levels out in the low-metallicity tail of the thin disk metallicity distribution (e.g., Santos et al. 2004). Right: the effect of total metallicity on giant planet occurrence. The hint of a plateau at low metallicity disappears when considering total metallicity. This occurs because non-solar abundance patterns (e.g., α-enhancement) begin to appear in stars with $[Fe/H] \approx -0.4$ as the thick disk stellar population starts to become considerable relative to the thin disk (e.g., Adibekyan et al. 2012a).

(A color version of this figure is available in the online journal.)

orbital period of $P_1 = 50$ days

$$a_1 = \left(\frac{P_1}{365}\right)^{2/3}, \tag{15}$$

$$A = a_1 + 10 \left(\frac{M_1}{3\,M_\odot}\right)^{1/3} a_1, \tag{16}$$

the smallest semimajor axis a_2 at which a $M_2 = 1\,M_{\mathrm{Jup}}$ giant planet would not make the system obviously unstable is

$$a_2 = A \left[1 - 10 \left(\frac{M_2}{3\,M_\odot}\right)^{1/3}\right]^{-1}. \tag{17}$$

I find that $a_2 \approx 1$ AU, or $P_2 \approx 365$ days. About 75% of the giant planets identified around solar-type stars with the RV technique have $P \geqslant 365$ days (Cumming et al. 2008), and therefore would not render a *Kepler* multiplanet system obviously unstable. For that reason, the fraction of the observed giant planet systems permitted by stability considerations is $\eta_{\mathrm{Hill}} = 0.75$. Schmitt et al. (2013), Cabrera et al. (2014), and Lissauer et al. (2014) recently identified KOI-351/Kepler-90 as a presumably stable system with multiple small, close-in planets with giant planets at longer orbital periods.

I use the completeness contours in Figure 6 of Mayor et al. (2011) to determine the expected number of long-period giant planets that would have been discovered around the RV multiplanet systems if every system had a long-period giant planet with mass and period drawn from the observed distributions of those quantities. Since all 20 RV multiplanet systems are in the catalog of Mayor et al. (2011), the completeness contours given in the paper apply for each system. For all p planets in the exoplanets.org catalog with $M \sin i > 100\,M_\oplus$ and $P > 365$ days, I check whether each is above the (100%, 95%, 80%, 60%, 40%) completeness contours. If so, I increment the expected number of detections q by (1, 0.95, 0.8, 0.6, 0.4). If the planet is below the 40% contour, I assume that it would be undetectable. I find that the fraction of giant planet systems that would have been recovered by Mayor et al. (2011) is $q/p = \eta_c = 0.71$. I illustrate this calculation in Figure 6.

3.3. Inference

The goal is to determine both the posterior distribution for P_{gp} inferred from the non-detection of giant planets in the RV multiplanet systems and the posterior of P_{gp} expected under the extrapolated MMEN scenario after taking into account stability and completeness. If there is tension between the two posteriors, then the MMEN scenario may not be an accurate description of *Kepler* multiplanet system formation.

Bayes' theorem guarantees

$$f(\theta|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\theta)f(\theta)}{\int f(\boldsymbol{y}|\theta)f(\theta)d\theta}, \tag{18}$$

where $f(\theta|\boldsymbol{y})$ is the posterior distribution of the model parameter θ, $f(\boldsymbol{y}|\theta)$ is the likelihood of the data $\boldsymbol{y}$ given θ, and $f(\theta)$ is the prior for θ. In this case, the likelihood is the binomial likelihood that describes the probability of a number of successes y in n Bernoulli trials each with probability θ of success:

$$f(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}. \tag{19}$$
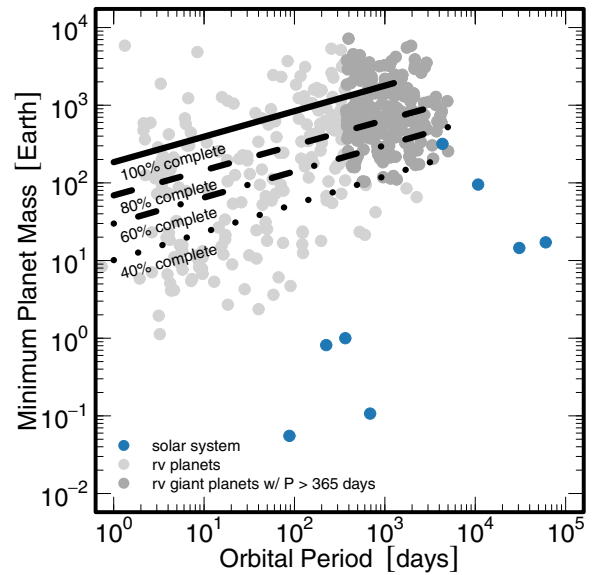
Figure 6. Distribution of planets discovered with the radial velocity technique in the semimajor axis–minimum mass plane. I plot average completeness curves from Mayor et al. (2011) indicating the fraction of planetary systems above each curve that would likely have been detected by HARPS+CORALIE. Close-in multiplanet systems with $P < 50$ days are likely only stable if there are no giant planets in those system with $P \lesssim 365$ days. Therefore, I only consider giant planets with orbital periods longer than 1 yr in my Monte Carlo calculations. Fully 75% of giant exoplanets are observed in this period range (e.g., Cumming et al. 2008), and I plot those planets included in my Monte Carlo in dark gray. The planets in the solar system are indicated as blue dots.

(A color version of this figure is available in the online journal.)

The calculation of the posterior $f(\theta|\boldsymbol{y})$ can be greatly simplified by the selection of an appropriate prior $f(\theta)$. In this case, it is possible to use a conjugate prior—a prior that guarantees that the posterior distribution will be in the same family as the prior. The Beta (α, β) distribution is a conjugate prior to the binomial likelihood and will give a Beta posterior, where α and β are the standard parameters of the Beta distribution. I take

$$f(\theta) = \mathrm{Beta}\,(\alpha, \beta), \tag{20}$$

$$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{0 \leqslant \theta \leqslant 1}, \tag{21}$$

where $\Gamma(x)$ is the standard gamma function and $I_{0 \leqslant \theta \leqslant 1}$ is the indicator function that is 1 in the interval [0, 1] and 0 elsewhere.

Plugging $f(y|\theta)$ and $f(\theta)$ into Bayes' theorem shows that the posterior distribution $f(\theta|y)$ can be written

$$f(\theta|y) = \frac{\Gamma(\alpha + y + \beta + n - y)}{\Gamma(\alpha + y)\Gamma(\beta + n - y)} \theta^{\alpha+y-1} (1-\theta)^{\beta+n-y-1} I_{0 \leqslant \theta \leqslant 1}, \tag{22}$$

$$= \mathrm{Beta}\,(\alpha + y, \beta + n - y). \tag{23}$$

In words, the posterior distribution of θ is itself a Beta function. The hyperparameters α and β of the prior can be thought of as encoding a certain amount of prior information in the form of pseudo-observations. Specifically, $\alpha - 1$ is the number of successes and $\beta - 1$ is the number of failures imagined to have already been observed and therefore included as prior information on θ. Taking any $\alpha = \beta = i$ where $i \geqslant 1$ could be thought of as an uninformative prior in the sense that the probability of success and failure in the prior distribution are

equally likely. However, if i is large then there is imagined to be a lot of prior information and the posterior distribution will mostly reflect the prior when $n \leqslant i$. On the other hand, if $n \gg i$, then the posterior will be dominated by the data. For that reason, I take $\alpha = \beta = 1$.

In the exoplanet context, θ is the unknown probability of giant planet occurrence P_{gp}, $n = 20$ is the number of RV multiplanet systems in Table 2, and y is the number of detected long-period giant planets. No giant planets have been detected in the RV multiplanet systems, so $y = 0$. After accounting for completeness and stability, the equivalent number of systems searched at 100% completeness would be $n' = \eta_c \eta_{Hill} n \approx (0.71)(0.75)(20) \approx 10.6$. The posterior distribution of P_{gp} inferred from the non-detection of long-period giant planets in the RV multiplanet systems is

$$P_{gp,obs} = \text{Beta}\,(\alpha + y, \beta + n' - y), \quad (24)$$

$$= \text{Beta}\,(1 + 0, 1 + \eta_c \eta_{Hill} n - 0), \quad (25)$$

$$= \text{Beta}\,(1, 11.6). \quad (26)$$

In the extrapolated MMEN scenario, every RV multiplanet system should have a long-period giant planet. Consequently, the expected number of successful planet discoveries after examining n systems should be $y' = \eta_c \eta_{Hill} n$. The posterior distribution of P_{gp} in the extrapolated MMEN scenario should be

$$P_{gp,MMEN} = \text{Beta}\,(\alpha + y', \beta + n - y'), \quad (27)$$

$$= \text{Beta}\,(1 + \eta_c \eta_{Hill} n, 1 + n - \eta_c \eta_{Hill} n), \quad (28)$$

$$= \text{Beta}(11.6, 10.4). \quad (29)$$

I plot these posterior distributions and 95% credible intervals in Figure 7. I find the probability that these posterior distributions overlap to be $p \approx 1 \times 10^{-3}$. This implies a 3σ (one-sided) difference between the two distributions: there are too few long-period giant planets observed in RV multiplanet systems for them to have formed in the extrapolated MMEN scenario. The concentration of solid material in the inner region of the disk hypothesized by Hansen & Murray (2012) should make long-period giant planets less common in the RV multiplanet systems than in the field planet host population. About 7% of field FGK stars have a giant planet with P in the range $1\,\text{yr} \leqslant P \leqslant 10\,\text{yr}$ (e.g., Cumming et al. 2008). Unfortunately, in RV multiplanet systems the current 1σ upper limit on long-period giant planet occurrence is 15%, so the Hansen & Murray (2012) scenario is unconstrained. In the future, a search for long-period giant planets in $n \approx 50$ RV multiplanet systems would be able to resolve the issue.

3.4. Metallicity Constraints

The frequency of giant planets can also be constrained by examining the metallicity distribution of a sample of stars. Giant planets at all orbital periods preferentially occur around metal-rich stars, so *Kepler* multiplanet systems that have a long-period giant planet should preferentially orbit metal-enriched stars. No strong metallicity effect has been measured for low-mass or small-radius planets orbiting solar-type stars, at least for the single systems that dominate the *Kepler* sample (e.g., Schlaufman & Laughlin 2011; Buchhave et al. 2012).
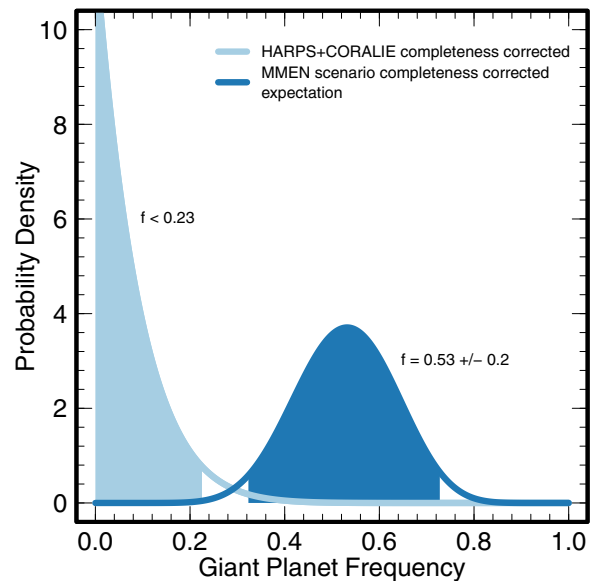

Figure 7. Posterior distribution of giant planet occurrence rate. The posterior distributions are given by the curves, while the 95% credible intervals are indicated by the shaded regions below the curves. The lack of announced giant planets in the 20 multiplanet systems discovered by HARPS combined with the completeness limits given in Mayor et al. (2011) indicate that the upper bound on the 95% credible interval on giant planet occurrence in those systems is 0.23. In the MMEN scenario, giant planets should be ubiquitous in these systems. The same completeness estimates indicate that the observed occurrence rate should be in the 95% credible interval 0.53 ± 0.20. The probability that the two distributions overlap is less than one in 1000.

(A color version of this figure is available in the online journal.)

To quantify the maximum fraction of *Kepler* multiplanet systems that could host a long-period giant planet yet not cause a noticeable metallicity enhancement in the *Kepler* multiplanet sample, I use a Monte Carlo simulation. I create many random mixed control samples composed of the metallicities of both multiple small-planet and large-planet host stars from Buchhave et al. (2012). I vary the fraction of large planet-host stars and then compare the metallicity distribution of the pure multiple small-planet host sample to the resultant mixed control sample distributions using the Anderson–Darling test. I plot the results in Figure 8 and find that control samples including less than about 50% large-planet host stars are generally consistent with the multiple small-planet host sample. Consequently, I expect that no more than 50% of the multiple small-planet host stars in the *Kepler* field also possess an unobserved long-period giant planet based on metallicity alone. This is in contrast to the extrapolated MMEN expectation, where nearly all systems of small planets should also have successfully formed a long-period giant planet companion. In fact, a giant planet host fraction of 1 suggested by the extrapolated MMEN scenario is rejected at the $p = 1 \times 10^{-3}$ level, or about 3σ (one-sided).

4. DISCUSSION

The MMEN scenario failed two independent tests of its apparent prediction that long-period giant planets should be ubiquitous in close-in multiplanet systems, each at 3σ. Failing these two independent tests at 3σ is approximately equivalent to failing a single test at 4σ. This tension indicates that massive disks alone cannot fully explain the properties of the *Kepler* multiplanet systems. If in situ formation and migration are the only two relevant processes, then migration must still have played a role. Raymond & Cossou (2014) reached a
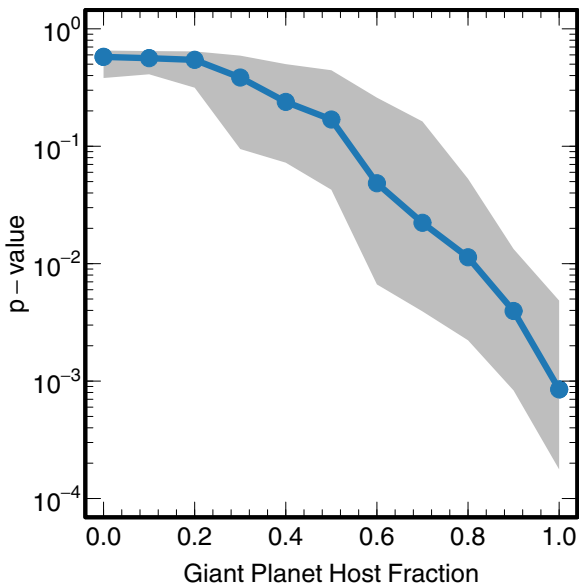
Figure 8. Anderson–Darling *p*-value as a function of the fraction of unobserved giant planet hosts in the sample of *Kepler* multiple small-planet hosts. The solid blue curve shows the median *p*-value after bootstrap resampling and the gray polygon shows the 1σ confidence interval. A giant planet host fraction of 1 suggested by the MMEN scenario is rejected at the $p = 1 \times 10^{-3}$ level, or about 3σ.

(A color version of this figure is available in the online journal.)

similar conclusion after examining disk surface density profiles. Therefore, the properties of the *Kepler* multiplanet systems can in principle be used to help determine the rate, direction, and stopping mechanism for migration in a gaseous disk as well as their dependencies on planet mass, stellar magnetic field, disk thermodynamics, turbulence, accretion rate, and dissipation mechanism.

It is possible that increasing the solid surface density of a disk by increasing metallicity does not have the same effect on giant planet formation as increasing the solid surface density through overall disk mass. It may be that the increased dust mass in a metal-rich protoplanetary disk efficiently removes ions and therefore affects the disk structure. The MRI-inactive regions of a metal-rich disk may therefore be larger than a solar-metallicity disk, and consequently promote especially efficient giant planet formation. In that case, the increased incidence of giant planets around metal-rich stars may not be a useful guide to giant planet formation in massive disks. It is also possible that the MMEN disk asymptotes to the MMSN disk beyond 1 AU, though in that case migration is still likely an important process in planet formation (e.g., Hansen & Murray 2012).

This study of in situ planet formation has produced many insights that are useful in the estimation of the frequency of exoplanet systems with "solar-system-like" architectures. Though this was not the aim of this effort, the topic is of considerably current interest and worth examining in detail. Indeed, the presence of long-period giant planets in systems of terrestrial planets may play an important role in the habitability of those planets. Wetherill (1994) argued that the formation of Jupiter prevented the formation of many comets and that Jupiter's subsequent presence ejected many more comets from the solar system. In systems with no long-period giant planets, the cometary impact flux in the terrestrial planet region is expected to be 1000 times the observed value in the solar system. Frequent comet impacts may sterilize a planet, and will

definitely inhibit the evolution of intelligent life. As a result, the presence of a long-period giant planet in an exoplanet system may play an important role in its habitability.

The fraction of solar-type stars that host a stable exoplanet system with at least one low-mass planet and at least one long-period giant planet is

$$\eta_{ss} = \eta_{\text{single}} \, \eta_{\text{sp}} \, \eta_{\text{gp}} \, \eta_{\text{Hill}}, \qquad (30)$$

where η_{single} is the fraction of solar-type stars in single star systems, η_{sp} is the fraction of solar-type stars that host small planets, η_{gp} is the fraction of solar-type stars that host a giant planet, and η_{Hill} is the fraction of systems that are Hill stable. The factor η_{Hill} is necessary because I assume that η_{sp} and η_{gp} are independent of each other and planet period. This assumption would lead to an overestimate of η_{ss}, as some of the hypothetical systems would be unstable. I correct after-the-fact using the factor η_{Hill} to ensure that I only count systems that would not be obviously unstable.

About 1/3 of the FGK stars in the solar neighborhood are in single systems, so $\eta_{\text{single}} = 0.33$ (e.g., Duquennoy & Mayor 1991). Fressin et al. (2013) report that 40% of solar-type stars have at least one planet with $P < 50$ days, so $\eta_{\text{sp}} = 0.4$. The factor η_{gp} depends sensitively on metallicity, so it is a function of the metallicity distribution of a stellar population. Because giant planet occurrence is such a strong function of metallicity, Galactic giant planets are only likely to be found in the Milky Way's disk or bulge. For the disk metallicity distribution, I use the distribution of [M/H] from Casagrande et al. (2011). For the bulge metallicity distribution, I use the distribution of [M/H] from Bensby et al. (2013), assuming that the abundances inferred from microlensed dwarf stars studied in that survey are representative of the bulge metallicity distribution. Also for the bulge sample, non-solar abundance patterns are important to consider. For that reason, I compute [M/H] from the measured values of [Fe/H], [O/Fe], [Mg/Fe], and [Si/Fe] assuming the solar abundances from Asplund et al. (2005). These four elements contribute 99% of the total stellar metallicity of the available abundances in the Bensby et al. (2013) catalog, and I therefore neglect the other elements. To calculate η_{gp} for each population, I use

$$\eta_{\text{gp}} = \frac{1}{m} \sum_{i=1}^{m} P_{\text{gp}}([\text{M/H}]_i), \qquad (31)$$

where m is the number of measured metallicities in a catalog and $P_{\text{gp}}([\text{M/H}]_i)$ is Equation (13). I find that $\eta_{\text{gp}} = 0.06$ for the disk and $\eta_{\text{gp}} = 0.17$ for the bulge.

Plugging in the numbers, I find that $\eta_{ss} = 0.0067$ in the disk and $\eta_{ss} = 0.017$ in the bulge. Recently, Petigura et al. (2013) published a tentative extrapolation of the frequency of Earth-size planets with orbital periods in the range $200\,\text{days} \leqslant P \leqslant 400\,\text{days}$; they found that $\eta_{\text{sp}} = \eta_{\oplus} \approx 0.06$. An Earth-mass planet orbiting at $a = 1$ AU implies that any giant planets in the system must be beyond $a \approx 3.5$ AU to avoid Hill instability. Using the giant planet orbital distribution from Cumming et al. (2008), that implies that $\eta_{\text{Hill}} = 0.44$. As a result, the fraction of solar-type stars hosting true solar system analogs including both an Earth-size planet near 1 AU and a long-period Jupiter-mass planet is $\eta_{ss} = 0.00059$ in the disk and $\eta_{ss} = 0.0015$ in the bulge.

The total number of such systems in the Galaxy is

$$N_{ss} = \eta_{\text{FGK}} \, (\eta_{ss,\text{disk}} \, M_{\text{disk}} + \eta_{ss,\text{bulge}} \, M_{\text{bulge}}), \qquad (32)$$

where $\eta_{\rm FGK}$ is the FGK mass fraction in a stellar population, $M_{\rm disk}$ is the stellar mass of the Milky Way's disk, and $M_{\rm bulge}$ is the stellar mass of the Milky Way's bulge. The stellar mass of the Milky Way's disk is $M_{\rm disk} \approx 4.6 \times 10^{10}\, M_\odot$, while the stellar mass of the bulge is $M_{\rm bulge} \sim 1 \times 10^{10}\, M_\odot$ (Bovy & Rix 2013; Widrow et al. 2008). I assume a Chabrier (2003) initial mass function (IMF) truncated at $0.08\, M_\odot$ and $120\, M_\odot$ for both the disk and bulge population, as there is no conclusive evidence that the IMF varies between the two populations (e.g., Bastian et al. 2010). Given the Chabrier (2003) IMF, I use the algorithm from the SLUGS code[8] to randomly sample the IMF and compute the fraction of stellar mass in FGK stars ($0.8\, M_\odot \leqslant M_* \leqslant 1.25\, M_\odot$). I find that 10% of a stellar population's stellar mass is in FGK stars, and because each star has $M_* \approx 1\, M_\odot$, the number of stars is equivalent to the stellar mass in FGK stars. As a result, $\eta_{\rm FGK} = 0.1$.

Putting all of the factors together, I find that the number of "solar systems" in the Galaxy with at least one Neptune-size planet with orbital period $P < 50$ days and a long-period giant planet is $N_{ss} \sim 5 \times 10^7$. The number of true solar system analogs in the Galaxy with an Earth-size planet in the habitable zone and a long-period Jupiter-mass giant planet is $N_{ss} \sim 4 \times 10^6$.

In this analysis, I have not accounted for the possibility of metallicity gradients in the disk of the Milky Way. The inner disk has a higher stellar density than the solar neighborhood, and it is likely more metal-rich too (e.g., Frinchaboy et al. 2013). Similarly, the outer disk has a lower stellar density than the solar neighborhood and may be metal-poor as well. This is a very active area of a research and future observations may resolve the issue.

5. CONCLUSIONS

Kepler multiple planet systems with several planets orbiting with periods $P < 50$ days may be difficult to explain in the traditional core accretion and Type I migration paradigm. In response, two in situ models of plant formation were proposed: the MMEN scenario of Chiang & Laughlin (2013) and the planetesimal migration scenario of Hansen & Murray (2012). Both models make predictions for the occurrence rate of long-period giant planets. In the extrapolated MMEN scenario, they are ubiquitous. In the planetesimal migration scenario, they are less common than in the field population. I find that the prediction from the extrapolated MMEN scenario for the occurrence of long-period giant planets in multiple low-mass systems discovered with the RV technique fails at 3σ. The lack of metallicity enhancement in the hosts of multiple small-planet systems discovered by *Kepler* provides an independent test of the extrapolated MMEN scenario, which it also fails at 3σ. I am unable to constrain the planetesimal migration scenario with the current sample. As a result, migration is still a necessary step in the formation of systems of close-in low-mass planets.

I also rederived the scaling of giant planet occurrence with metallicity, and I find that $P_{\rm gp} = 0.05^{+0.02}_{-0.02} \times 10^{(2.1\pm0.4)[{\rm M/H}]} = 0.08^{+0.02}_{-0.03} \times 10^{(2.3\pm0.4)[{\rm Fe/H}]}$. I used these relations to calculate the frequency of "solar systems" in the Galaxy, where a solar system is defined as a single FGK star orbited by a planetary system with at least one small planet interior to the orbit of a giant planet. The presence of a giant planet exterior to an Earth-size planet may be necessary to prevent frequent comet impacts from inhibiting the evolution of life on an otherwise

habitable planet. I find that in the solar neighborhood, about 0.7% of solar-type stars have a "solar system" consisting of both a Neptune-size planet with $P < 50$ days and a protective long-period companion. Intriguingly, I find that true solar system analogs with both a terrestrial planet in the habitable zone and a long-period giant planet companion to protect it occur around only 0.06% of solar-type stars. There are perhaps 4×10^6 such systems in the Galaxy.

APPENDIX

KEPLER MULTIPLANET SYSTEM QUERY

The following SQL query can be used in the *Kepler* CasJobs database available from MAST to reproduce my sample of *Kepler* multiple small-planet candidate systems. It is first necessary to set "Context" to "kepler."

```
SELECT a.kepid, a.kepoi_name, a.koi_prad,
 a.koi_period, a.koi_srad,
c.kepler_name
FROM kepler_koi a
INNER JOIN (
    SELECT kepid
    FROM kepler_koi
    WHERE koi_disposition != 'FALSE POSITIVE'
    AND koi_prad < 5
    GROUP BY kepid
    HAVING COUNT(kepoi_name) > 1
) b ON a.kepid = b.kepid
LEFT OUTER JOIN published_planets c
ON a.kepid = c.kepid AND a.kepoi_name =
 c.kepoi_name
INNER JOIN keplerObjectSearchWithColors d ON
 a.kepid = d.kic_kepler_id
WHERE a.koi_disposition != 'FALSE POSITIVE'
AND d.jh BETWEEN 0.22 AND 0.62
AND d.hk BETWEEN 0.00 AND 0.10
AND d.jk BETWEEN 0.22 AND 0.72
ORDER BY a.kepoi_name, a.koi_period;
```

REFERENCES

Adibekyan, V. Z., Santos, N. C., Sousa, S. G., et al. 2012a, A&A, 543, A89
Adibekyan, V. Z., Sousa, S. G., Santos, N. C., et al. 2012b, A&A, 545, A32
Alibert, Y., Baraffe, I., Benz, W., et al. 2006, A&A, 455, L25

[8] At https://sites.google.com/site/runslug/ and described in Fumagalli et al. (2011) and da Silva et al. (2012).

Andrews, S. M., Rosenfeld, K. A., Kraus, A. L., & Wilner, D. J. 2013, ApJ, 771, 129
Asplund, M., Grevesse, N., & Sauval, A. J. 2005, in ASP Conf. Ser. 336, Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis, ed. T. G. Barnes, III & F. N. Bash (San Francisco, CA: ASP), 25
Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339
Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24
Bensby, T., Yee, J. C., Feltzing, S., et al. 2013, A&A, 549, A147
Binney, J., & Merrifield, M. 1998, Galactic Astronomy (Princeton, NJ: Princeton Univ. Press)
Bonfils, X., Delfosse, X., Udry, S., et al. 2013, A&A, 549, A109
Borucki, W. J., Koch, D. G., Basri, G., et al. 2011a, ApJ, 728, 117
Borucki, W. J., Koch, D. G., Basri, G., et al. 2011b, ApJ, 736, 19
Bovy, J., & Rix, H.-W. 2013, ApJ, 779, 115
Buchhave, L. A., Latham, D. W., Johansen, A., et al. 2012, Natur, 486, 375
Burke, C. J., Bryson, S. T., Mullally, F., et al. 2014, ApJS, 210, 19
Butler, R. P., Johnson, J. A., Marcy, G. W., et al. 2006, PASP, 118, 1685
Cabrera, J., Csizmadia, S., Lehmann, H., et al. 2014, ApJ, 781, 18
Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, A&A, 530, A138
Chabrier, G. 2003, PASP, 115, 763
Chambers, J. E., Wetherill, G. W., & Boss, A. P. 1996, Icar, 119, 261
Chatterjee, S., Hadi, A. S., & Price, B. 2000, Regression Analysis by Example (3rd ed.; New York: Wiley)
Chiang, E., & Laughlin, G. 2013, MNRAS, 431, 3444
Chiang, E., & Youdin, A. N. 2010, ARE&PS, 38, 493
Conover, W. J. 1999, Practical Nonparametric Statistics (3rd ed.; New York: Wiley)
Covey, K. R., Ivezić, Ž., Schlegel, D., et al. 2007, AJ, 134, 2398
Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, PASP, 120, 531
da Silva, R. L., Fumagalli, M., & Krumholz, M. 2012, ApJ, 745, 145
Duquennoy, A., & Mayor, M. 1991, A&A, 248, 485
Endl, M., Cochran, W. D., Kürster, M., et al. 2006, ApJ, 649, 436
Fabrycky, D. C., Lissauer, J. J., Ragozzine, D., et al. 2012, arXiv:1202.6328
Fischer, D. A., & Valenti, J. 2005, ApJ, 622, 1102
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Frinchaboy, P. M., Thompson, B., Jackson, K. M., et al. 2013, ApJL, 777, L1
Fumagalli, M., da Silva, R., Krumholz, M., & Bigiel, F. 2011, in ASP Conf. Ser. 440, UP2010: Have Observations Revealed a Variable Upper End of the Initial Mass Function?, ed. M. Treyer, T. K. Wyder, J. D. Neill, M. Seibert, & J. C. Lee (San Francisco, CA: ASP), 155
Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279
Hansen, B. M. S., & Murray, N. 2012, ApJ, 751, 158
Hansen, B. M. S., & Murray, N. 2013, ApJ, 775, 53
Hayashi, C. 1981, PThPS, 70, 35
Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27

Hou, A., Parker, L. C., Harris, W. E., & Wilman, D. J. 2009, ApJ, 702, 1199
Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2005, Icar, 179, 415
Ida, S., & Lin, D. N. C. 2005, ApJ, 626, 1045
Kley, W., & Nelson, R. P. 2012, ARA&A, 50, 211
Laughlin, G., Bodenheimer, P., & Adams, F. C. 2004, ApJL, 612, L73
Lissauer, J. J., Marcy, G. W., Bryson, S. T., et al. 2014, ApJ, 784, 44
Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., et al. 2011, ApJS, 197, 8
Lo Curto, G., Mayor, M., Benz, W., et al. 2013, A&A, 551, A59
Lovis, C., Mayor, M., Pepe, F., et al. 2006, Natur, 441, 305
Lovis, C., Ségransan, D., Mayor, M., et al. 2011, A&A, 528, A112
Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, ApJS, 210, 20
Mayor, M., Marmier, M., Lovis, C., et al. 2011, arXiv:1109.2497
Mayor, M., Udry, S., Lovis, C., et al. 2009, A&A, 493, 639
Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23
Paardekooper, S.-J., Baruteau, C., Crida, A., & Kley, W. 2010, MNRAS, 401, 1950
Pepe, F., Lovis, C., Ségransan, D., et al. 2011, A&A, 534, A58
Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, PNAS, 110, 19273
R Core Team 2013, R: A Language and Environment for Statistical Computing (Vienna: R Foundation for Statistical Computing)
Raymond, S. N., & Cossou, C. 2014, MNRAS, 440, L11
Rivera, E. J., Butler, R. P., Vogt, S. S., et al. 2010, ApJ, 708, 1492
Santos, N. C., Israelian, G., & Mayor, M. 2004, A&A, 415, 1153
Schlaufman, K. C., & Laughlin, G. 2011, ApJ, 738, 177
Schmitt, J. R., Wang, J., Fischer, D. A., et al. 2013, arXiv:1310.5912
Schneider, J., Dedieu, C., Le Sidaner, P., Savalle, R., & Zolotukhin, I. 2011, A&A, 532, A79
Smith, A. W., & Lissauer, J. J. 2009, Icar, 201, 381
Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M., & Udry, S. 2011a, A&A, 533, A141
Sousa, S. G., Santos, N. C., Israelian, G., et al. 2011b, A&A, 526, A99
Sousa, S. G., Santos, N. C., Mayor, M., et al. 2008, A&A, 487, 373
Stephens, M. A. 1974, J. Amer. Stat. Assn., 69, 730
Taylor, M. B. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
Valenti, J. A., & Fischer, D. A. 2005, ApJS, 159, 141
van Leeuwen, F. 2007, Hipparcos, The New Reduction of the Raw Data (Astrophysics and Space Science Library, Vol. 350; Berlin: Springer)
Vogt, S. S., Wittenmyer, R. A., Butler, R. P., et al. 2010, ApJ, 708, 1366
Weidenschilling, S. J. 1977, Ap&SS, 51, 153
Weiss, L. M., & Marcy, G. W. 2014, ApJL, 783, L6
Wetherill, G. W. 1994, Ap&SS, 212, 23
Widrow, L. M., Pym, B., & Dubinski, J. 2008, ApJ, 679, 1239
Wolfgang, A., & Laughlin, G. 2012, ApJ, 750, 148
Wright, J. T., Fakhouri, O., Marcy, G. W., et al. 2011, PASP, 123, 412
Wu, Y., & Lithwick, Y. 2013, ApJ, 772, 74