The Astronomical Journal, 159:188 (10pp), 2020 May https://doi.org/10.3847/1538-3881/ab7814

On the Patterns Observed in *Kepler* Multi-planet Systems

Wei Zhu (祝伟)
Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada; weizhu@cita.utoronto.ca
Received 2019 April 14; revised 2020 February 17; accepted 2020 February 18; published 2020 April 7

Abstract

Recent studies claimed that planets around the same star have similar sizes and masses and regular spacings, and that planet pairs usually show ordered sizes such that the outer planet is usually the larger one. Here I show that these patterns can be largely explained by detection biases. The *Kepler* planet detections are set by the transit signal-to-noise ratio (S/N). For different stellar properties and orbital period values, the same S/N corresponds to different planetary sizes. This variation in the detection threshold naturally leads to apparent correlations in planet sizes and the observed size ordering. The apparently correlated spacings, measured in period ratios, between adjacent planet pairs in systems with at least three detected planets are partially due to the arbitrary upper limit that the earlier study imposed on the period ratio, and partially due to the varying stability threshold for different planets. After these detection biases are taken into account, we do not find strong evidence for the so-called intra-system uniformity or the size ordering effect. Instead, the physical properties of *Kepler* planets are largely independent of the properties of their siblings and the parent star. It is likely that the dynamical evolution has erased the memory of *Kepler* planets about their initial formation conditions. In other words, it will be difficult to infer the initial conditions from the observed properties and the architecture of *Kepler* planets.

Unified Astronomy Thesaurus concepts: Exoplanets (498); Exoplanet systems (484); Astrostatistics strategies (1885)

1. Introduction

Almost all planets reside in multi-planet systems, and the physical and orbital properties of planets in the same system convey important clues about their formation and evolution. To date the majority of the multi-planet systems were found by *Kepler* (Borucki et al. 2010).[1] However, *Kepler* could only detect those planets that transited their hosts and had transit signals above some certain noise level. This limits our knowledge about the detected multi-planet systems and complicates the theoretical interpretation.

As *Kepler* observations provide directly the transit planet-to-star radius ratio, the relative sizes of planets inside the same system can be easily examined. Based on data from the first four months of observations, Lissauer et al. (2011) pointed out that adjacent planets were likely to have very similar radii, as most of the pairs show $R_{\rm p,in}/R_{\rm p,out} \approx 1$. Here $R_{\rm p,in}$ and $R_{\rm p,out}$ are the radii of the inner and the outer transiting planets (tranets), respectively. This feature was further studied in Ciardi et al. (2013), and the authors reported that most (>60%) of the multi-planet systems found by *Kepler* appeared to have this size–location correlation: the outer planet was larger than the inner one.[2] In both studies, the authors compared the observed and the simulated radius ratio distributions to determine the statistical significance of their finding. Their simulated distributions were produced by randomly drawing radii from the *observed* radius distribution and going through customized signal-to-noise ratio (S/N) cuts. As I will explain later, this approach does not capture all the detection biases.

Later follow-up efforts that lead to better characterizations of *Kepler* stars, allowing for comparisons of planetary parameters across systems. Recently, Weiss et al. (2018b, hereafter W18) claims that planetary systems are like "peas in a pod," namely that the planets orbiting around the same host have similar sizes and regular spacings. They took the large sample of *Kepler* multi-planet systems whose parameters were refined by the California-*Kepler* Survey (CKS; Petigura et al. 2017; Johnson et al. 2017), sorted the *CKS planets*[3] in the same systems according to their orbital periods, and computed the correlation between sizes of neighboring CKS planets. They then quantified the significance of this correlation through bootstrap tests and found that the observed correlation could not be explained by randomly resampling the observed size distribution. The procedure was similar for the spacings between planets.

A later[4] work by Millholland et al. (2017) adopted a similar statistical approach and further claimed that the masses of planets inside the same system, given by Hadden & Lithwick (2017) from analyzing the transit timing variations (TTVs; Agol et al. 2005; Holman & Murray 2005), should also be similar. This, together with the aforementioned trends about radius and spacing, was summarized as the intra-system uniformity (Millholland et al. 2017).

However, an issue that was overlooked in these studies (Lissauer et al. 2011; Ciardi et al. 2013; Millholland et al. 2017; W18) is the detection threshold. Below I use the transit detection as an example, but the idea applies to TTV mass measurements as well (see Section 7). The *Kepler* transit search pipeline requires a nominal minimum S/N of 7.1; for the planet

[1] See a list of all known multi-planet systems at the NASA Exoplanet Archive (http://exoplanetarchive.ipac.caltech.edu; Akeson et al. 2013).

[2] Ciardi et al. (2013) argued that this result only applied to planet pairs in which at least one was approximately Neptune-sized or larger. However, without this constraint the results are qualitatively similar in the statistical sense. See their Figures 4 and 11.

[3] A CKS planet is a planet that transits the host, is detectable by *Kepler*, and is included in the CKS sample. CKS planets are almost certainly valid planets, but because of the geometric transit probability and *Kepler* detection sensitivity, the CKS planets are not necessarily all the planets in those systems.

[4] The W18 work was posted on the arXiv pre-print server before the Millholland et al. (2017) work.

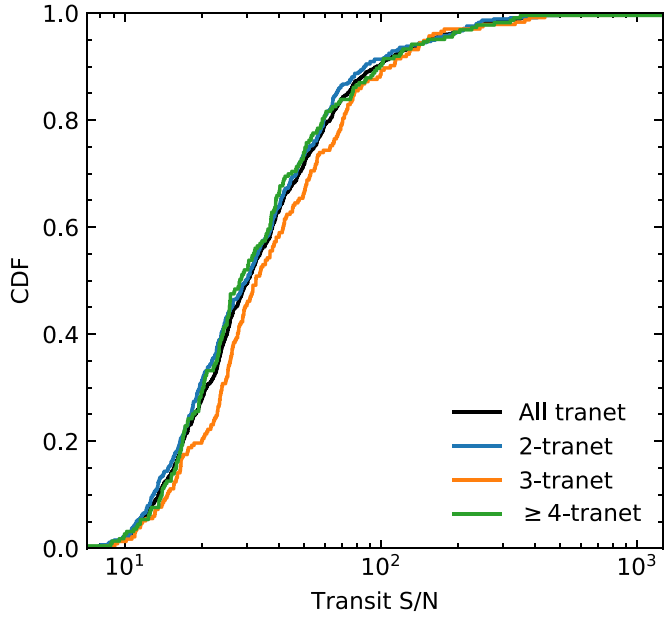

Figure 1. Cumulative distributions of transit S/N. Different transit multiplicities are shown in different colors, and the overall sample is shown in black. There is no clear dependence of S/N on the transit multiplicity.

sample used in W18, a higher (S/N = 10) threshold was used. The detection completeness depends strongly on the S/N at the low end (e.g., Fressin et al. 2013; Thompson et al. 2018). Given the dominating contribution from smaller planets and weaker transits (e.g., Hsu et al. 2019), the S/N values of *Kepler* transit detections as a result appear to pile up toward the detection threshold (e.g., Figure 1 of Ciardi et al. 2013) and are not affected by the variations of stellar parameters or noise levels. For a certain S/N threshold, the resulting planet radius threshold depends on the orbital period as well as the host properties. This variation was not fully taken into account by those referenced works in generating simulated parameters.

A more proper way is fully forward modeling the detection and selection processes from the *intrinsic* planetary (radius or mass) distribution. Instead of randomly drawing parameters from the observed distribution, one should draw from the intrinsic distribution and then apply the same detection criteria (e.g., S/N cut) on these simulated planets. This process requires knowing the intrinsic planet distribution function and having access to the automated *Kepler* detection pipeline. It is further complicated by the fact that the planet distribution function is period dependent (e.g., Dong & Zhu 2013; Hsu et al. 2019) and possibly multiplicity dependent, and that the *Kepler* detection efficiency is weakly multiplicity dependent (Zink et al. 2019).

There is a shortcut that circumvents these problems. In the full forward modeling approach, one generates synthetic planetary systems, adds stellar noises, passes them to the *Kepler* detection pipeline, decides which planets are detectable, and finally performs statistical analyses on the simulated detections. Through this whole process planetary physical parameters are converted into transit observables and the detectability of an individual planet is controlled by the transit S/N. Given the central role of transit S/N, we can directly start from this parameter as a shortcut to the full forward modeling approach. As shown in Figure 1 and the upper right panel of Figure 2, the S/N distribution is independent of the transit multiplicity and the stellar properties. Therefore, we can randomly draw detections from this universal S/N distribution, derive planetary parameters, and perform the same statistical analysis as we do on the real data. As the relation between S/N and planetary radius indicates (Equation (1)), when the minor contribution from the orbital period is ignored, correlated sizes will definitely lead to correlated S/N values, but correlated S/N values do not necessarily mean correlated sizes because of the same stellar size and stellar noise level that the two adjacent planets share. Therefore, by randomly sampling the S/N distribution and thus assuming no correlation in S/N values, I am being more generous than just assuming no size correlation.

In this work, I apply this more robust statistical approach to study the patterns observed in *Kepler* multi-planet systems. I describe the planet sample in Section 2 and explain the basic idea in Section 3. Then in Sections 4 and 5 I discuss the issues involved in the "peas in a pod" claim. The size–location correlation is re-evaluated in Section 6. Finally, I briefly comment on the intra-system mass uniformity claim and then discuss the results in Section 7.

2. Sample

I use the same multi-planet sample as in W18. This sample includes 909 CKS planets in 355 multi-planet systems. The parameters of the individual planets and of their hosts were provided in Table 1 of W18. Of relevance to this study are the stellar mass $M_\star$, stellar radius $R_\star$, 6 hr combined differential photometric precision ($\mathrm{CDPP}_{6\,\mathrm{hr}}$, a measure of the stellar noise level; Christiansen et al. 2012), impact parameter b, planetary radius R_p, and orbital period P. This table did not include a column of S/N, but it can be easily computed by

$$\mathrm{S/N} = \frac{(R_\mathrm{p}/R_\oplus)^2\sqrt{3.5\ \mathrm{yr}/P}}{\mathrm{CDPP}_{6\,\mathrm{hr}}\sqrt{6\ \mathrm{hr}/T}}. \quad (1)$$

Here T is the transit duration, given by

$$T = 13\ \mathrm{hr}\left(\frac{P}{1\ \mathrm{yr}}\right)^{1/3}\left(\frac{\rho_\star}{\rho_\odot}\right)^{-1/3}\sqrt{1-b^2}, \quad (2)$$

where $\rho_\star$ is the stellar mean density. Note that the Equation (2) of W18 did not have the factor $\sqrt{1-b^2}$, which made their S/N overestimated, although this would only have a minor effect on the results.

I show in Figure 2 the stellar noise level $\mathrm{CDPP}_{6\,\mathrm{hr}}$ versus three chosen parameters: planet radius R_p, period ratio of adjacent CKS planets, and the transit detection S/N. For demonstration purposes I divide the whole sample into two at the median $\mathrm{CDPP}_{6\,\mathrm{hr}}$: the *quiet sample* and the *noisy sample*.

The S/N value dictates the significance of a transit detection and is a more fundamental observable in signal detections than planetary radius, which is not even a direct observable in transit light curves. Because S/N already takes into account the variation of stellar noise level $\mathrm{CDPP}_{6\,\mathrm{hr}}$ (Equation (1)), one does not expect the S/N distribution to be different between the quiet and the noisy samples. Indeed, a two-sample Kolmogorov–Smirnov (KS) test between the S/N distributions from two samples gives $p = 0.18$,[5] confirming that the S/N distribution is invariant to the variations of stellar properties.

[5] The Anderson–Darling test, which is more sensitive to the differences in scales as well as the tails of the distributions, gives $p = 0.05$.

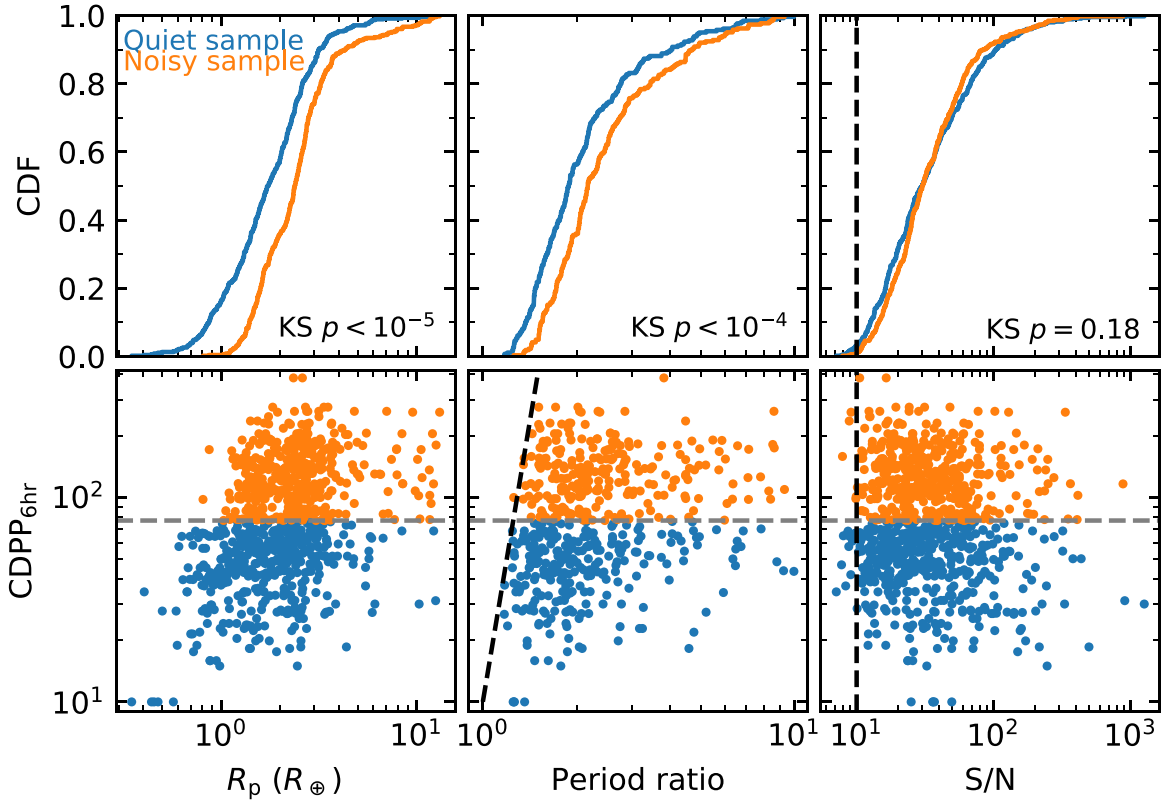

Figure 2. Lower panels show the noisy level of the planet host, characterized by $\mathrm{CDPP}_{6\,\mathrm{hr}}$, vs. the planet radius R_p (lower left), planetary period ratio (lower middle), and signal-to-noise ratio (S/N) of the transit detection (lower right). The gray dashed horizontal lines mark the median $\mathrm{CDPP}_{6\,\mathrm{hr}}$, based on which the sample is divided into two. In the upper panels we compare the cumulative distribution functions (CDF) of the individual parameter (R_p, period ratio, and S/N) from the two samples, and the two-sample KS test p-values are indicated. The radius and period ratio distributions of the two samples are statistically different. In the case of period ratio (lower middle panel), we also mark with the black dashed line the smallest $\mathrm{CDPP}_{6\,\mathrm{hr}}$ for various period ratios. The S/N distributions are very similar between the two samples. Nearly 50% of planet detections have S/N < 30, i.e., only a factor of three above the detection threshold (S/N $= 10$, as marked by the vertical dashed line).

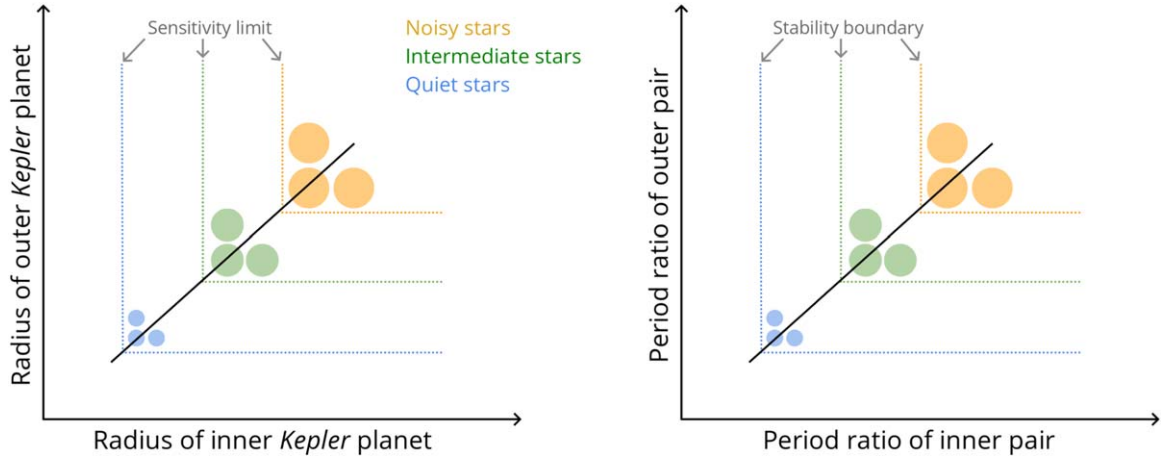

Figure 3. Schematic view of how an inhomogeneous stellar sample will lead to the size (left panel) and spacing (right panel) correlations. For stars with a certain noise level, only planets above a certain size are detectable. This can be seen in the lower left panel of Figure 2. Given that there are more smaller planets than larger ones, detected planet pairs tend to cluster around the corner that is defined by the vertical and horizontal sensitivity limits. For different stellar samples (as measured by their noise levels), these clusterings appear at different locations along the diagonal line, and thus a collection of these planets will naturally show a size correlation (left panel). Similarly, the stability boundary in the measure of period ratio decreases with decreasing planet sizes, as seen in the lower middle panel of Figure 2. Because of the strong bias in the orbital period of transit detections, observable planet triplets will tend to cluster around the corner of the stability boundary. For different stellar samples (as measured by the noise level), these clusterings appear at different locations along the diagonal line, and thus a collection of these planet triplets will naturally show a spacing correlation (right panel).

However, the radius distributions and the period ratio distributions from the two samples are different: the two-sample KS test gives $p < 10^{-5}$ and $p < 10^{-4}$, respectively. These differences are most prominent at small values of R_p and period ratio. The difference in R_p is due to the projection of the same S/N distribution into different stellar samples: smaller planets are more easily detected around more quiet stars. Then through the dynamical stability requirement, the difference in R_p distributions propagates into the difference in period ratio distributions. See Section 5 for more details.

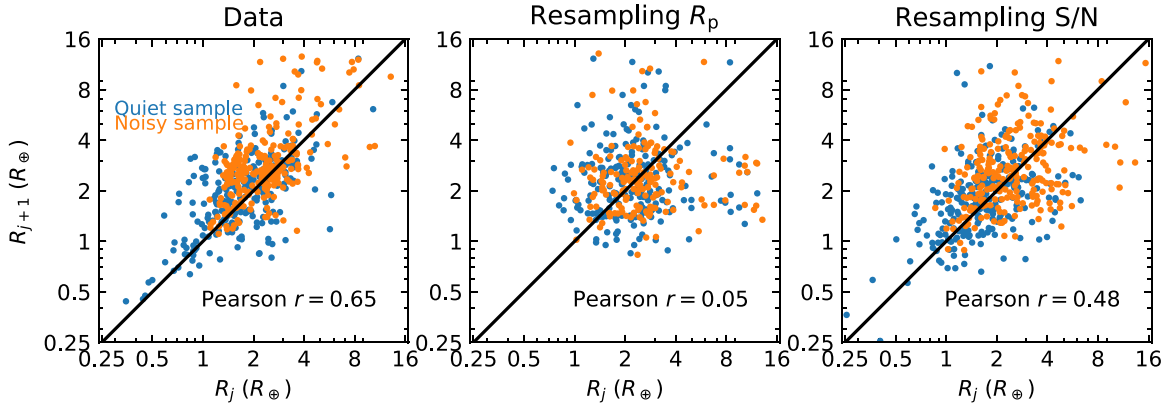

Figure 4. Comparisons between radii of adjacent planets (R_j and R_{j+1}). In the left panel is the distribution of planet pairs in the CKS multi-planet sample. In the middle and the right panels are the synthetic planet pairs generated in two different approaches: resampling R_p and resampling S/N. The former was used in W18. Planet pairs from two stellar samples (quiet and noisy) are plotted with different colors to highlight the difference. The correlation coefficient between $\log R_j$ and $\log R_{j+1}$ is also indicated at the lower right corner of each panel.

Therefore, across the whole sample there is a universal S/N distribution but no universal radius distribution or period ratio distribution. The latter two were used in the bootstrap test by W18.

3. The Idea of this Paper

Figure 3 illustrates how the variation in the detection threshold originating from a fixed S/N can lead to size and spacing correlations in observed planets.

First of all, one should know that there are more smaller planets than larger ones, at least down to *Kepler*'s sensitivity limit, after the correction of detection bias (e.g., Hsu et al. 2019). Therefore, if only planets above a certain size are detectable, then the detected planets will tend to pile up toward the detection threshold. In realistic missions such as *Kepler*, the detection threshold is usually fixed in S/N because of its central role in signal detection. However, given the relation between planetary radius and transit S/N (Equation (1)) a fixed S/N threshold will lead to different radius thresholds for stars with different noise levels. Such a varying radius threshold will naturally lead to a correlation between the sizes of neighboring *Kepler* planets. See the left panel of Figure 3 for a simple illustration.

W18 used bootstrap tests on planetary radii to study the significance of the size correlation. The underlying assumption behind their radius bootstrap test is that the radius distribution, $P(R_p)$, should be universal across different stellar subsamples. I have shown in the previous section that this assumption does not hold for the sample under investigation. Instead, the radius distribution depends on the stellar properties, specifically the noise level $\mathrm{CDPP}_{6\,\mathrm{hr}}$. I denote this conditional radius distribution as $P(R_p|\mathrm{CDPP}_{6\,\mathrm{hr}})$ and note that

$$P(R_p) \neq P(R_p|\mathrm{CDPP}_{6\,\mathrm{hr}}). \quad (3)$$

A proper radius bootstrap test should therefore be randomly drawing radii from the conditional radius distribution $P(R_p|\mathrm{CDPP}_{6\,\mathrm{hr}})$. Unfortunately, such a conditional radius distribution cannot be easily specified, but one can use the relation between R_p and S/N to further simplify the procedure. With other parameters the same (as is required by the bootstrap test), R_p uniquely determines S/N. Thus randomly sampling $P(R_p|\mathrm{CDPP}_{6\,\mathrm{hr}})$ is equivalent as randomly sampling S/N from $P(\mathrm{S/N}|\mathrm{CDPP}_{6\,\mathrm{hr}})$ and then deriving R_p from S/N. Recalling that the $\mathrm{CDPP}_{6\,\mathrm{hr}}$ distribution is universal across different stellar subsamples,

$$P(\mathrm{S/N}|\mathrm{CDPP}_{6\,\mathrm{hr}}) = P(\mathrm{S/N}), \quad (4)$$

one can further simplify the proper radius bootstrap test as randomly resampling the S/N distribution $P(\mathrm{S/N})$. This is another justification of the S/N-resampling method that is used in this work.

Similar to the claimed size correlation, the claimed spacing correlation is also affected by the variation in the detection threshold. With decreasing stellar noise, the minimum period ratio between adjacent planets also decreases, as shown in the lower middle panel of Figure 2. This is probably because of the stability boundary, in the measure of period ratio, decreases with decreasing planet size and we refer to Section 5 for a more detailed explanation. As the transit probability strongly biases toward small period values and thus small period ratios, the detectable planet pairs tend to pile up toward the smallest period ratio (i.e., the stability boundary). The variation of this stability boundary in an inhomogeneous stellar sample naturally leads to a spacing correlation. See the right panel of Figure 3 for an illustration.

4. On the Radius Uniformity

In the left panel of Figure 4 I show the radius of one CKS planet, R_j, versus the radius of the outer adjacent CKS planet, R_{j+1}. This is very similar to Figure 2 of W18. The correlation coefficient, quantified by Pearson r, between $\log R_j$ and $\log R_{j+1}$ of all planet pairs is $r = 0.65$, consistent with the value reported in W18.[6]

To assess the importance of this correlation W18 generated synthetic planet systems in which the radius of each planet was randomly drawn from the overall radius distribution. An example realization following the procedure of W18 is shown in the middle panel of Figure 4. Note that in this plot there appears to be fewer sub-Earth-sized planets. This is because,

[6] As pointed out by the journal's statistics editor, Pearson's r correlation coefficient assumes linear relationships and Gaussian scatter, neither of which are well established for this data set. Nonparametric correlation coefficients, such as Kendall's τ, are better suited for this correlation test. Pearson's r values are nevertheless used in this work simply to keep consistency with W18.

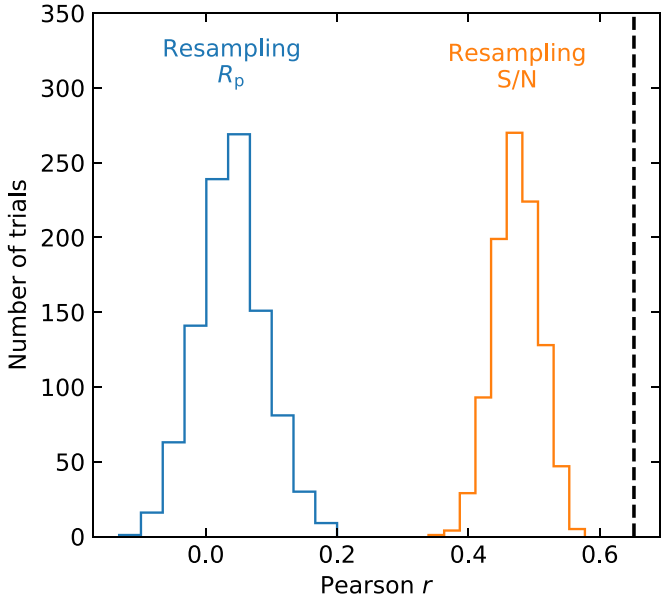

Figure 5. Distributions of the Pearson r coefficients from statistical tests. The blue and orange histograms are results from two different approaches, and the black dashed line indicates the measured correlation value ($r = 0.65$). Note that resampling radius is a bootstrap method whereas resampling S/N is a forward modeling approach.

following W18, simulated planets with S/N < 10 are excluded. This step reduces the number of planets by nearly 20%.

However, as discussed in Section 2 and illustrated in Figure 2, the S/N distribution is more fundamental and universal than the radius distribution in transit signal detections. In particular, the radius distribution appears differently for stars with different noise levels. This can also be seen in the left panel of Figure 4, where I have differentiated the planet pairs from quiet and noisy samples. Note the similarity between this plot and the left panel of Figure 3. Following the reasoning in Section 3, I therefore modify the bootstrap test of W18. Instead of resampling the R_p distribution, I resample the S/N distribution and then, with other parameters unchanged, derive R_p from Equation (1).[7] Note that while I am bootstrapping transit S/N, I am essentially performing a forward modeling (see Section 1 for the detailed explanation). The result from one random test is shown in the right panel of Figure 4. In this simulated sample the planet pairs from two stellar samples show a systematic offset, a feature that is similar to the data (left panel). The radii of adjacent planets also show significant correlation, $r = 0.48$.

I repeat the above statistical test for 1000 times, record all Pearson r coefficients, and show their histogram in Figure 5. For comparison purposes I also produce the histogram of r coefficients from 1000 bootstrap tests following the W18 procedure (i.e., resampling R_p), and the resulting histogram peaks at $r \approx 0$, similar to what W18 had (see their Figure 5). By resampling on the more fundamental parameter S/N, I almost always reproduce, at least qualitatively, the observed size correlation, although with an average correlation coefficient of $r \approx 0.5$ this effect alone cannot explain quantitatively the observed size correlation. We discuss below what can potentially account for the remaining size correlation.

[7] In practice, this can be easily achieved with the relation $R_p^{new} = R_p\sqrt{(S/N)_{new}/(S/N)}$.

Since *Kepler* can only detect transiting planets above a certain S/N threshold, it is very likely that many of the *Kepler* multi-planet systems may contain additional undetectable planets. Outside the period limit that *Kepler* can probe ($\sim$1 yr), studies have shown that cold giant planets preferentially co-exist with inner small planets (Zhu & Wu 2018; Bryan et al. 2019; Herman et al. 2019), the inclusion of which will certainly break the size similarity pattern.

Inside the *Kepler* period domain, there are also signs of additional planets. Dynamical studies have suggested that the majority of the *Kepler* multi-planet systems are not fully packed if the detected planets are all the planets in the system. According to Fang (2013) $\sim$55% of systems with at least four detected planets can contain additional intervening planets without leading to dynamical instability, and the fraction is even higher for systems with two or three detected planets (see also Pu & Wu 2015). Furthermore, the fact that *Kepler* planet detections pile up toward the detection threshold[8] is also suggesting that smaller and undetectable planets do exist.

An intervening undetectable planet between two detectable ones likely also transits and has a smaller (compared to the detection threshold) size. What the addition of such a smaller intervening planet does to the radius correlation (e.g., left panel of Figure 4) is two-fold. First, one planet pair that shows strong correlation is removed and then, two planet pairs that show much weaker correlation are added. The combined consequence is that the size correlation is reduced significantly. To demonstrate this effect, one would like to increase *Kepler*'s sensitivity to recover the smaller planets. This is obviously not practical, so I turn to the opposite direction.[9] I increase the S/N threshold used in the statistical test, which is equivalent to increasing stellar noises and thus lowering *Kepler*'s sensitivity, and then measure the size correlation in the same way. As shown in Figure 6, the size correlation becomes stronger in such down-graded *Kepler* missions. This is aligned with our speculation and suggests that, in a superior *Kepler* mission which can detect much smaller planets and thus is less affected by detection biases, the size correlation should be much weaker.

Another way to show the influence of additional planets on the size correlation is to restrict to high-multiple systems that are less likely to contain additional planets because of stability requirements. I only include systems with at least four CKS planets and repeat the same statistical tests. The results are shown in Figure 7. This time the distribution of the correlation coefficients from statistical tests is statistically closer to the observed value. This again confirms that the missing planets do have an effect on the size correlation.

5. On the Period Ratio Uniformity

W18 also claimed that the spacings between planets, measured by the period ratios, are correlated in systems with at least three CKS planets. To reach this conclusion, they first identified all CKS planet triplets, which consisted of all the CKS planets in three-planet systems and three consecutive planets in higher-multiple systems. For each planet triplet they

[8] The detection efficiency of the *Kepler* pipeline depends on the transit S/N. As Ciardi et al. (2013) have shown with some earlier *Kepler* samples (see their Figure 1), Kepler detection is only complete for S/N $\gtrsim$ 25. Therefore, when the incompleteness of the detection pipeline is taken into account, S/N of $\sim$20 is still at the edge of the detection threshold.

[9] This test was originally suggested by Xi Zhang.

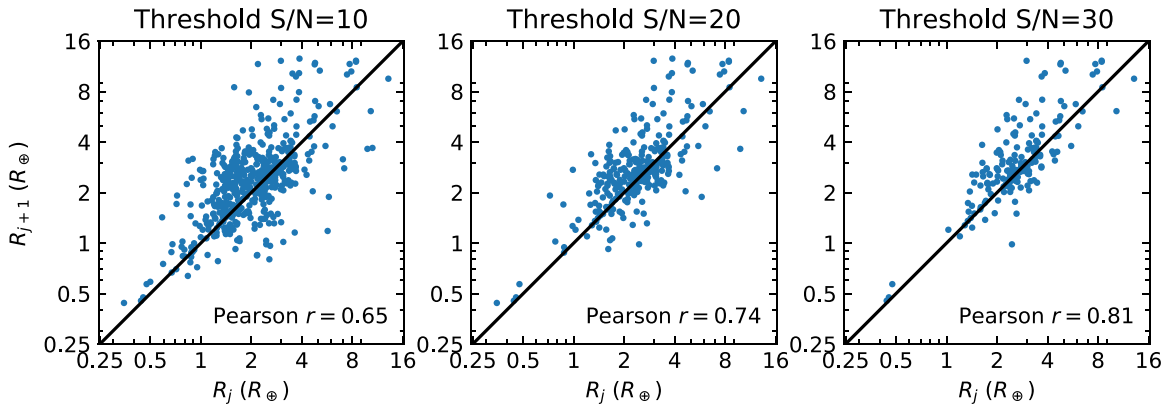

Figure 6. Size correlation plots for different choices of the threshold S/N values. Increasing the threshold S/N is equivalent to decreasing the *Kepler* sensitivity, which leads to increased size correlation. Extrapolating to lower S/N thresholds, this test suggests that a much better *Kepler*-like mission that is sensitive to smaller planets will find a much weaker size correlation.

computed the period ratio between the inner two planets, $\mathcal{P}_{\rm in}$, and the period ratio between the outer two planets, $\mathcal{P}_{\rm out}$. Considering the incomplete sensitivity to large period ratios, W18 only included the planet triplets whose $\mathcal{P}_{\rm in} < 4$ and $\mathcal{P}_{\rm out} < 4$. Then they computed the Pearson r coefficient between the two variables $\log \mathcal{P}_{\rm in}$ and $\log \mathcal{P}_{\rm out}$ and found $r = 0.46$. To assess the significance of this correlation, they generated synthetic systems, in which the period ratios were randomly drawn from the overall period ratio distribution, and found that the correlations in these simulated samples were systematically much smaller than the observed one.

This approach may produce biased results in two ways. First, the cut at period ratio $\mathcal{P} = 4$ is fairly arbitrary and not physically motivated. In the left panel of Figure 8 I show the $\mathcal{P}_{\rm out}$ versus $\mathcal{P}_{\rm in}$ for all planet triplets in the W18 sample. One can see that the detection limit at large period ratios is diagonal rather than flat. This can be well explained by the detectability of these multi-planet systems. If the planets in the same system have coplanar or nearly coplanar orbits, the detectability of all planets only depends on the orbital period of the outermost planet, $P_{\rm outermost}$. For a planet triplet, this detection threshold scales as $\propto \mathcal{P}_{\rm in}\mathcal{P}_{\rm out}$. In the left panel of Figure 8, I plot the line that corresponds to $\mathcal{P}_{\rm in}\mathcal{P}_{\rm out} = 25$, and it roughly agrees with the upper boundary of all data points. Adopting this physically motivated detection threshold, I find the Pearson $r = 0.21$ between $\log \mathcal{P}_{\rm in}$ and $\log \mathcal{P}_{\rm out}$. Restricting to $\mathcal{P}_{\rm in}\mathcal{P}_{\rm out} < 16$ gives $r = 0.15$. Both correlations are much weaker than that given by W18.

The second issue is the varying detection threshold of the transit. At first glance, the S/N, as given in Equation (1), has only weak dependence on orbital period and no explicit dependence on the period ratio. The period ratio comes into play via the dynamical stability requirement. The stability boundary is typically measured in the number (K) of mutual Hill radii, $r_{\rm H}$,

$$a_2 - a_1 = K \cdot r_{\rm H}, \quad r_{\rm H} \equiv \frac{a_1 + a_2}{2}\left(\frac{m_1 + m_2}{3}\right)^{1/3}, \qquad (5)$$

where a_i and m_i are the semimajor axis and mass of the inner ($i = 1$) and outer ($i = 2$) planets, respectively. For simplicity, we further assume $m_1 \approx m_2 \approx M_\oplus (R_{\rm p}/R_\oplus)^3$. Note that this is not valid in general, but it is acceptable for the planet pairs that are just above the detection threshold and close to the instability limit. Then with Kepler's third law we can have a rough scaling between the planetary size and the critical period ratio for dynamical stability

$$\mathcal{P} = \frac{P_2}{P_1} \approx 1 + 0.019K\left(\frac{R_{\rm p}}{R_\oplus}\right). \qquad (6)$$

Below we adopt $K = 20$, although in reality the threshold on K also depends on many factors, such as individual planet masses, eccentricities, mutual inclinations, etc. (e.g., Chambers et al. 1996; Zhou et al. 2007; see Pu & Wu 2015 for a detailed discussion). As Figure 2 shows, the smallest planet detectable around a typical noisy star is $1R_\oplus$, for which the stability threshold is $\mathcal{P}_{\rm crit,1} \approx 1.4$. The smallest detectable around a typical quiet star, in contrast, is $0.5R_\oplus$, with a stability threshold of $\mathcal{P}_{\rm crit,2} \approx 1.2$. This varying threshold is visible in the lower middle panel of Figure 2 as well as the right panel of Figure 8. Again for demonstration purposes I have differentiated the planet triplets from the quiet sample and noisy sample with different colors. No planet triplets from the noisy sample are below the red dashed line, which denotes $\mathcal{P}_{\rm in}\mathcal{P}_{\rm out} = 2 \approx \mathcal{P}^2_{\rm crit,1}$, whereas planet triplets from the quiet sample can extend further down to $\mathcal{P}^2_{\rm crit,2}$. This varying stability threshold was not taken into account in W18.

The varying stability threshold applying to an inhomogeneous stellar sample naturally leads to a spacing correlation, as is illustrated in the right panel of Figure 3. Note the similarity between this plot and the right panel of Figure 8. Generating synthetic planetary systems that meet all detection thresholds of individual planets and of the triplet as well as the stability threshold is not trivial, so I cannot assess quantitatively the impact of this effect on the Pearson r coefficient. However, as a qualitative check, if only planet triplets from the noisy sample are used, I have $r = 0.25$ even with the square cut at $\mathcal{P} = 4$. This is much smaller than what one has ($r = 0.46$) if both noisy and quiet samples are used.

6. On the Size Ordering

Lissauer et al. (2011) and Ciardi et al. (2013) first noticed that the *Kepler* multi-planet systems show a size–location correlation. Specifically, the larger planet in any planet pair is most often the one with the longer period. To check the statistical significance against observation biases, these authors compared the radius ratio distributions between observation

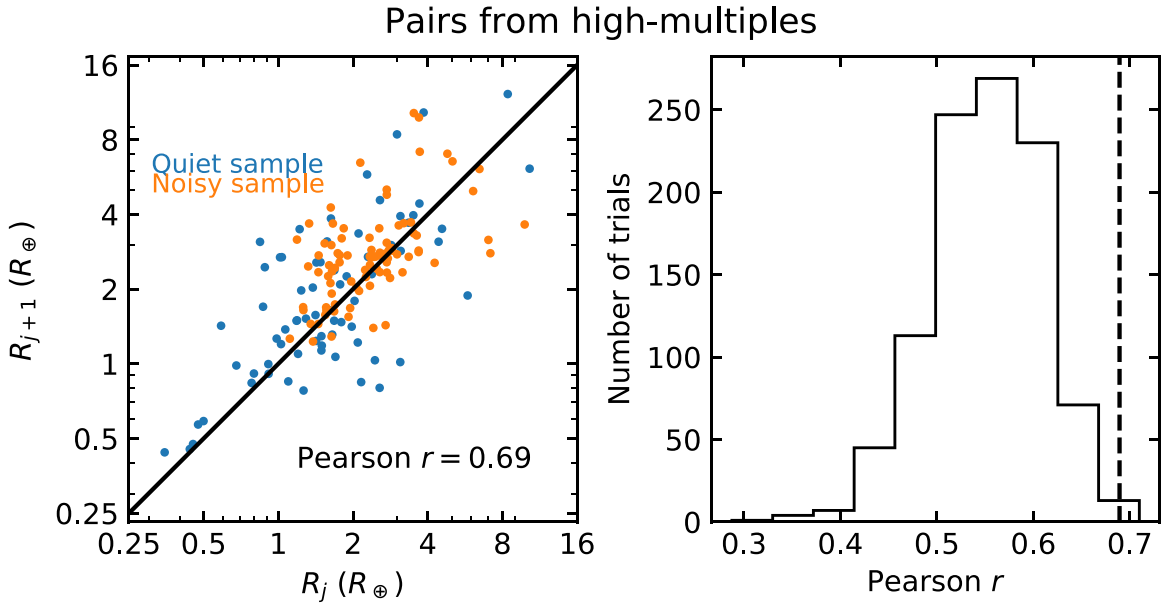

Figure 7. The left panel shows the comparison between radii of adjacent planets (R_j and R_{j+1}) from compact systems, defined as systems with at least four transiting planets. The planet pairs from two stellar samples are marked with different colors. The correlation coefficient is given in the lower right corner. For these planet pairs, 1000 statistical tests are performed, in which the S/N (rather than R_p) is resampled, and the resulting Pearson r values are shown as the histogram in the right panel. The observed correlation coefficient is not much different from the correlation we get in the statistical test.

and simulation. In generating simulated planet pairs, they randomly drew radii from the observed radius distribution and then, to mimic their selection procedure, eliminated those which would not be detected if either of the planets at the orbital period of the other one fell below the specified S/N threshold. Their simulated radius ratio distribution showed an equal number of planet pairs with $R_{p,in} < R_{p,out}$ and $R_{p,in} > R_{p,out}$. Ciardi et al. (2013) also performed several other tests, including using different S/N thresholds and maximum periods. The size ordering was always observed. Therefore, it was concluded that the size–location correlation has a physical origin.

This statistical approach suffers from the same issue as the W18 one. Using the CKS multi-planet sample, I show in Figure 9 the cumulative distributions of transit S/Ns and radius ratios between planets in pairs. Similar to what Lissauer et al. (2011) and Ciardi et al. (2013) found, here I also have more than 60% of planet pairs showing the so-called size–location correlation: $R_{p,in} < R_{p,out}$. However, the S/N is on average unity, suggesting that the transit signal of the inner planet is as strong as that of the outer one. This is is expected if one randomly pairs up the transit S/N values from the observed S/N distribution. See the green curve in the left panel of Figure 9. Given the relation between S/N and radius (Equation (1)), one can derive the radius ratio from the S/N,

$$\frac{(S/N)_{in}}{(S/N)_{out}} \approx \left(\frac{R_{p,in}}{R_{p,out}}\right)^2 \left(\frac{P_{in}}{P_{out}}\right)^{-1/3}. \quad (7)$$

In the above approximation we ignore the contribution from impact parameters. Because of the term involving the period ratio, two transit signals with equal S/N naturally lead to a pair of planets with $R_{p,in} < R_{p,out}$, that is, the size–location correlation.

The distributions from randomly sampling the S/N distribution do not match the observed distributions perfectly, in particular in the range $1 < (S/N)_{in}/(S/N)_{out} < 4$, or equivalently $1 < R_{p,in}/R_{p,out} < 2$. There are two possible reasons. First, transiting planets in some pairs do show a weak size correlation. This is also suggested in Section 4, as the random S/N sampling cannot fully explain the observed correlation strength (Figure 5). However, as is also suggested in Section 4, this can possibly be explained by transit bias, as one is not detecting all of the planets in the same system. Regardless, the fraction of planet pairs that show the size correlation is likely a small fraction ($\lesssim$5%). Otherwise it would require a very fine-tuned period ratio distribution to push the median S/N to almost exactly unity.

Another possible reason that can account for the deviation is some subtle detection bias in the planet search pipeline. When S/N values are randomly paired up, it is implicitly assumed that the detection efficiency does not depend on parameters other than S/N. This is not entirely true in reality. For the same value of S/N, the detection efficiency decreases gradually with the orbital period (see, e.g., Figure 9 of Thompson et al. 2018). This effect biases against planet pairs with large $R_{p,in}/R_{p,out}$, the type of pairs that are in short for the observational distribution to match the simulated one. Future detailed studies are needed to quantify this effect.

In short, the observed size–location correlation in *Kepler* multi-planet systems can be mostly, if not fully, explained by detection biases. It is possible that some planet pairs do have ordered sizes, but they only consist of a small fraction ($\lesssim$5%) of all planet pairs.

7. Discussion

In this work, I re-examine several claims about the relative sizes and spacings between *Kepler* planets around the same host. I make use of the observed transit S/N values, because they are observationally more fundamental than other parameters such as the planet radius. I present several findings.

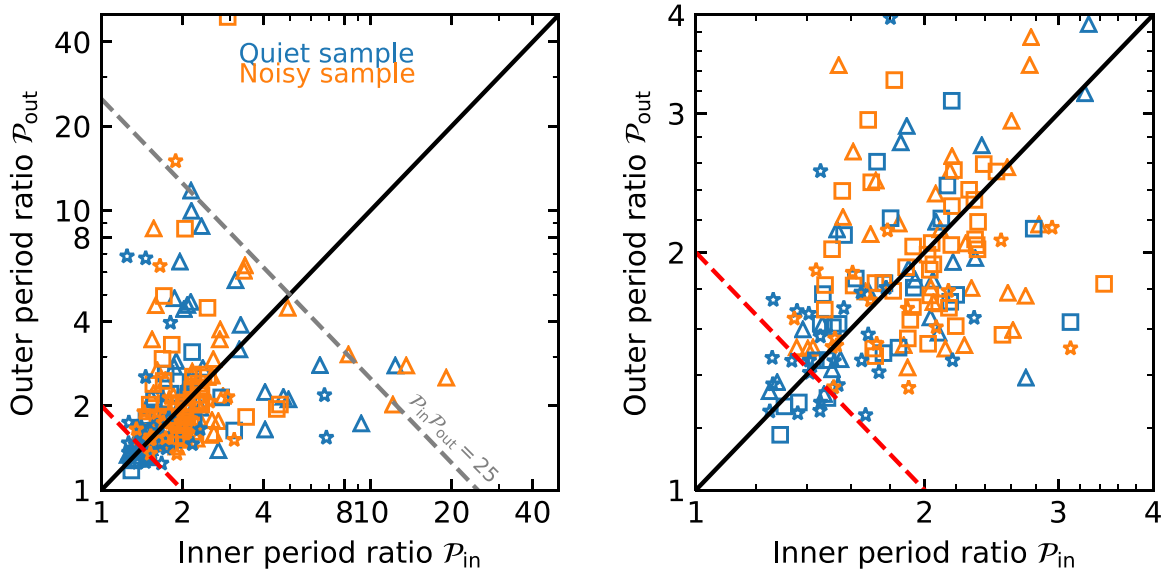

Figure 8. Outer period ratio vs. the inner period ratio for all triplets (left panel) and the subset with both period ratios <4 (right panel). I use triangles, squares, and asterisks for planet triplet from three-tranet, four-tranet, and ⩾five-tranet systems. Weiss et al. (2018b) only used planet triplets with both period ratios <4 (i.e., those in the right panel). However, such a square cut is not physically motivated, since the detection probability of a triplet scales as $\mathcal{P}_{in}\mathcal{P}_{out}$. The gray dashed line in the left panel shows an example. I also label with different colors the planet triplets from two stellar samples to highlight their different distributions. In particular, there is no planet triplet with $\mathcal{P}_{in}\mathcal{P}_{out} < 2$ (i.e., to the lower left of the red dashed lines).

1. The apparently similar sizes of planets in the same *Kepler* system can be largely explained by the projection of the same S/N cut onto different stellar properties.
2. The apparently correlated spacings, measured in period ratios, between adjacent planet pairs in systems with at least three detected planets are partially due to the arbitrary upper limit that W18 imposed on the period ratio and partially due to the varying stability threshold in different stellar samples.
3. The observed size–location correlation can be explained by the projection of the same S/N onto different values of orbital period. As far as the transit detection is concerned, the inner and the outer transiting planets on average have similar S/N values.

The claim of intra-system mass uniformity by Millholland et al. (2017) suffers from very similar issues. Below I draw the analogy between the Millholland et al. (2017) and W18 studies and defer a more quantitative analysis for future works. The analysis of Millholland et al. (2017) was performed on a sample of 89 planets from 37 *Kepler* systems, whose masses were constrained by Hadden & Lithwick (2017) through TTV.[10] Whether or not a TTV mass measurement can be made is more directly related to the TTV amplitude than to the planet mass. For a pair of planets, the TTV amplitude is generally dependent on the distance from the period commensurability, Δ, which is defined as (Lithwick et al. 2012)

$$\Delta \equiv \left| \frac{P_{out}}{P_{in}} \frac{J'}{J} - 1 \right| . \quad (8)$$

[10] Millholland et al. (2017) had eight planets with only mass upper limits, as indicated in Hadden & Lithwick (2017), in their sample: Kepler-23 d, Kepler-24 e, KOI-115.03, Kepler-105 b, Kepler-114 b, Kepler-114 d, Kepler-310 c, and KOI-427.01. It is not appropriate to treat mass upper limits and mass measurements in the same way. The removal of these planets reduces the number of transiting planets in four systems (Kepler-105, 114, 310, 549) down to one, so these systems are excluded from the sample. In the end I have 77 planets from 33 systems.

Here P_{in} and P_{out} are the orbital periods of the inner and outer planet in a TTV pair, respectively, and J'/J is the closest small integer ratio for P_{in}/P_{out}. Other things being equal, a smaller Δ means that a lower planet mass can be measured from TTV. This makes the $\Delta - m_p$ relation (Figure 10) somewhat analogous to the $CDPP_{6\,hr}$–R_p relation (lower left panel of Figure 2). Consequently, Millholland et al. (2017) reshuffling the planet mass is similar to W18 bootstrapping the planet radius. As shown in Section 4, this approach leads to biased results.

Therefore, the so-called intra-system uniformity and the size ordering effect that appear in the *Kepler* multi-planet systems can be mostly, if not entirely, explained by observational biases. As far as the data is able to inform, the physical properties of one *Kepler* planet are largely independent of the properties of both its siblings and the parent star.

So far the analysis has been done on the *Kepler* multi-planet systems, but the same conclusion likely applies to all *Kepler* planets. It is true that over half of the transiting planets were found in systems with only one transiting planet (i.e., single-tranet systems). However, this is most likely a result of a selection effect, as *Kepler* only detects planets that transit the host star. The orbital properties, such as eccentricity and mutual inclination, of single tranets and multi tranets are different (e.g., Xie et al. 2016; Zhu et al. 2018; Van Eylen et al. 2019), but this does not necessarily mean that their physical properties are different as well. In fact, studies have shown that the planetary properties and the stellar properties of single-tranet and multi-tranet systems are similar (e.g., Munoz Romero & Kempton 2018; Weiss et al. 2018a; Zhu et al. 2018), suggesting they are likely the same population. Nevertheless, future ground-based radial velocity observations will be able to tell whether or not this is true.

A recent work by He et al. (2019) applied the full forward modeling method to study the *Kepler* multi-planet systems. The authors generated multi-planet systems following specific

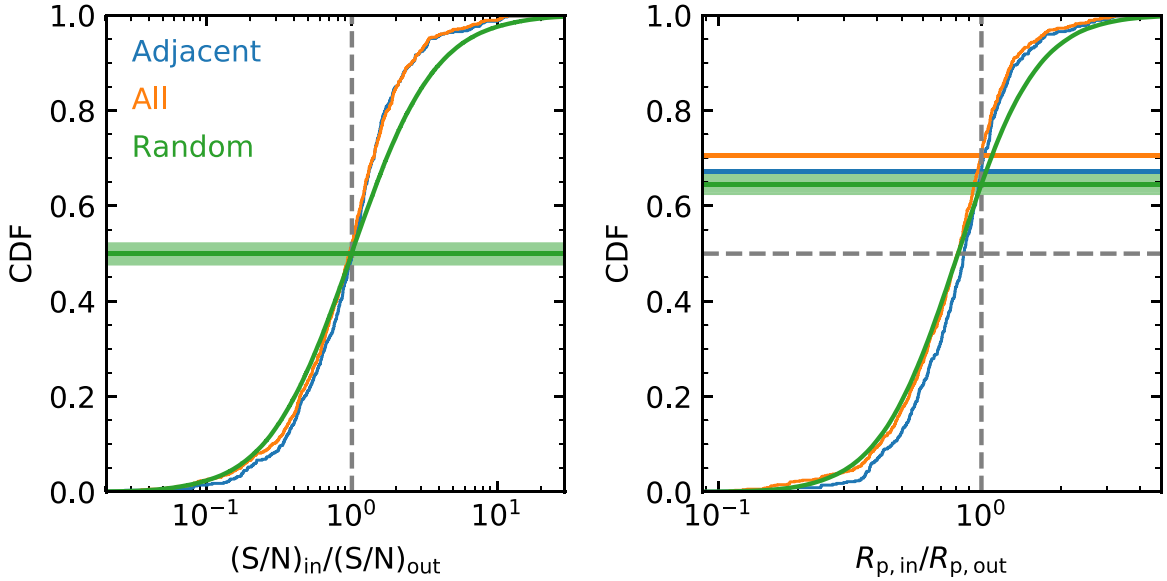

Figure 9. Cumulative distributions of the transit S/N (left panel) and the radius ratio (right panel) between planets in a pair. The ratio is specified as the property of the inner to the property of the outer. The two ways of forming planet pairs are studied. The blue curves use only pairs from adjacent planets, whereas the orange curves include pairs from non-adjacent planets. The green curves are results from randomly sampling the S/N distribution. The vertical dashed lines mark the boundary where the quantity of the inner equals the quantity of the outer, and the horizontal solid lines with colors mark the values at which the curves meet these equalities. The green regions mark the 1σ confidence interval, derived from 1000 realizations of the random sampling, and the horizontal dashed lines mark the median. On average the inner tranet has the same S/N as the outer one, which is what one expects from the random sampling. This naturally leads to an excess of pairs with $R_{p,in} < R_{p,out}$ (i.e., the size–location correlation).

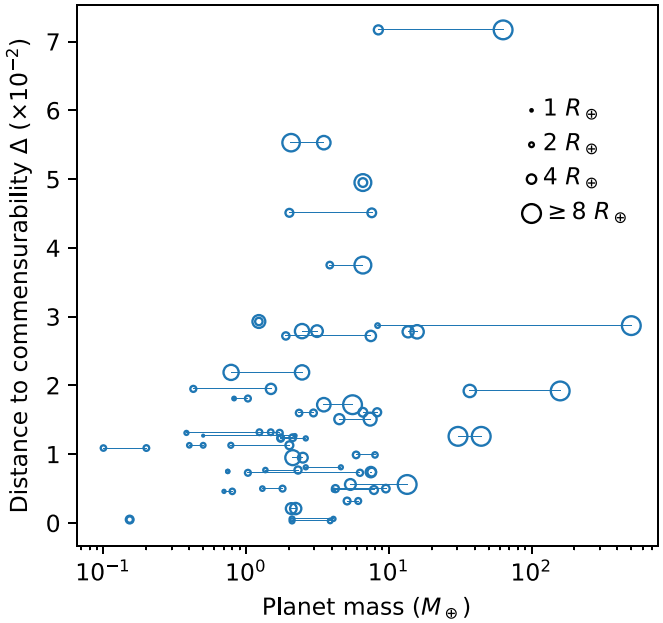

Figure 10. Illustration of the planet pairs used in Millholland et al. (2017) for claiming the intra-system mass uniformity. For each pair, I show on the x-axis the masses of individual planets and on the y-axis the distance to exact period commensurability, Δ, given by Table 2 of Hadden & Lithwick (2017). The size of the symbol reflects the radius of the planet. As Δ decreases, lower masses can be detected via the TTV technique. This detection bias is not taken into account by Millholland et al. (2017) in constructing synthetic systems.

prescriptions, passed the simulated systems through a simplified *Kepler* detection pipeline, and compared their simulated planet catalogs to the real catalog. The authors find a better match in a combination of selected observables between the simulated and real catalogs once the periods and radii of planets around the same host are assumed to be correlated. However, it is unclear whether the improvement in the fit is due to physical correlations or some artifacts in the model. In fact, as their best-fit models show (Figures 3–5 of He et al. 2019), the match to transit depth and transit depth ratio distributions, the most relevant ones for the size correlation, is not improved once the size correlation is introduced.[11] This is in agreement with the conclusion of the current paper that there is no evidence for the size correlation. The spacing correlation is a much more complicated issue in such a full forward modeling approach. To give a specific example, how to generate stable multi-planet systems remains an unsolved problem. The critical spacing for long-term stability depends on many factors, including the number of planets (i.e., multiplicity; Funk et al. 2010) and orbital properties (i.e., eccentricity and mutual inclination; Pu & Wu 2015), the latter of which have also been shown to be multiplicity dependent (Xie et al. 2016; Zhu et al. 2018). Using a fixed K value for all systems, as was done in He et al. (2019), is not realistic. More work is needed.

The conclusion that the properties of *Kepler* planets are largely independent of the properties of their siblings and the parent star has theoretical implications. Either the formation of *Kepler* planets had almost no requirement for their birth environment, or the (likely) chaotic evolution erased their memory of the initial condition. This latter scenario is more likely once several other pieces of evidence are put together. *Kepler* multi-planet systems are shown to be dynamically compact (Pu & Wu 2015) and have very diverse compositions (Wu & Lithwick 2013; Marcy et al. 2014; Hadden & Lithwick 2017). The nearly flat period ratio distribution, arisen from either in situ formation (Petrovich et al. 2013; Wu et al. 2019) or breaking the chain of resonances after migration (Izidoro et al. 2017), also points to a stage of dynamical instability. Finally, *Kepler* planets are shown to be strongly

[11] Their best-fit clustered periods and sizes model actually gives a worse match to the observed distributions of transit depth and transit depth ratios than their best-fit non-clustered model does.

correlated with outer giant planets (Zhu & Wu 2018; Bryan et al. 2019; Herman et al. 2019). The planet–planet scatterings that are responsible for the large eccentricities of the cold giant planets (Chatterjee et al. 2008; Jurić & Tremaine 2008) can easily drive dynamical instabilities in the inner system.

As the orbital velocity far exceeds the escape velocity for the majority of *Kepler* planets, the encounters between planets during the dynamical evolution can significantly revise their physical properties, thus removing the imprint of the initial formation conditions. In other words, it will be difficult to infer the initial conditions from the properties of current *Kepler* planets. A different conclusion was reached in Kipping (2018) from studying the size orderings of *Kepler* planets. Kipping (2018) assumed that the observed orderings are physical and free from observational biases. This is not true, as the present study has shown.

I would like to thank Subo Dong, Cristobal Petrovich, Yanqin Wu, Norm Murray, and Eve Lee for discussions, and particularly Xi Zhang for the suggestion about the new test in the size correlation section. I also thank the anonymous referees for the comments and suggestions on the manuscript. I also thank Lauren Weiss for kindly sharing her code that helps reproduce their results. W.Z. was supported by the Beatrice and Vincent Tremaine Fellowship at CITA.

ORCID iDs

Wei Zhu (祝伟) https://orcid.org/0000-0003-4027-4711

References

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567
Akeson, R. L., Chen, X., Ciardi, D., et al. 2013, PASP, 125, 989
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Sci, 327, 977
Bryan, M. L., Knutson, H. A., Lee, E. J., et al. 2019, AJ, 157, 52
Chambers, J. E., Wetherill, G. W., & Boss, A. P. 1996, Icar, 119, 261
Chatterjee, S., Ford, E. B., Matsumura, S., et al. 2008, ApJ, 686, 580
Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279
Ciardi, D. R., Fabrycky, D. C., Ford, E. B., et al. 2013, ApJ, 763, 41
Dong, S., & Zhu, Z. 2013, ApJ, 778, 53
Fang, J., & Margot, J.-L. 2013, ApJ, 767, 115
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Funk, B., Wuchterl, G., Schwarz, R., et al. 2010, A&A, 516, A82
Hadden, S., & Lithwick, Y. 2017, AJ, 154, 5
He, M. Y., Ford, E. B., & Ragozzine, D. 2019, MNRAS, 490, 4575
Herman, M. K., Zhu, W., & Wu, Y. 2019, AJ, 157, 248
Holman, M. J., & Murray, N. W. 2005, Sci, 307, 1288
Hsu, D. C., Ford, E. B., Ragozzine, D., et al. 2019, AJ, 158, 109
Izidoro, A., Ogihara, M., Raymond, S. N., et al. 2017, MNRAS, 470, 1750
Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, AJ, 154, 108
Jurić, M., & Tremaine, S. 2008, ApJ, 686, 603
Kipping, D. 2018, MNRAS, 473, 784
Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., et al. 2011, ApJS, 197, 8
Lithwick, Y., Xie, J., & Wu, Y. 2012, ApJ, 761, 122
Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, ApJS, 210, 20
Millholland, S., Wang, S., & Laughlin, G. 2017, ApJL, 849, L33
Munoz Romero, C. E., & Kempton, E. M.-R. 2018, AJ, 155, 134
Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017, AJ, 154, 107
Petrovich, C., Malhotra, R., & Tremaine, S. 2013, ApJ, 770, 24
Pu, B., & Wu, Y. 2015, ApJ, 807, 44
Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, ApJS, 235, 38
Van Eylen, V., Albrecht, S., Huang, X., et al. 2019, AJ, 157, 61
Weiss, L. M., Isaacson, H. T., Marcy, G. W., et al. 2018a, AJ, 156, 254
Weiss, L. M., Marcy, G. W., Petigura, E. A., et al. 2018b, AJ, 155, 48
Wu, D.-H., Zhang, R. C., Zhou, J.-L., et al. 2019, MNRAS, 484, 1538
Wu, Y., & Lithwick, Y. 2013, ApJ, 772, 74
Xie, J.-W., Dong, S., Zhu, Z., et al. 2016, PNAS, 113, 11431
Zhou, J.-L., Lin, D. N. C., & Sun, Y.-S. 2007, ApJ, 666, 423
Zhu, W., Petrovich, C., Wu, Y., Dong, S., & Xie, J. 2018, ApJ, 860, 101
Zhu, W., & Wu, Y. 2018, AJ, 156, 92
Zink, J. K., Christiansen, J. L., & Hansen, B. M. S. 2019, MNRAS, 483, 4479