



# Predicting the Dominant Formation Mechanism of Multiplanetary Systems

Cheyenne Shariat<sup>1,2</sup>, Yasuhiro Hasegawa<sup>2</sup>, Bradley M. S. Hansen<sup>1</sup>, Tze Yeung Mathew Yu<sup>1</sup>, and Renyu Hu<sup>2,3</sup><sup>1</sup>Department of Physics and Astronomy, University of California, Los Angeles, Los Angeles, CA 90095, USA; [cheyanneshariat@ucla.edu](mailto:cheyanneshariat@ucla.edu)<sup>2</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA<sup>3</sup>Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA

Received 2023 December 20; revised 2024 January 31; accepted 2024 February 6; published 2024 March 15

## Abstract

Most, if not all, Sun-like stars host one or more planets, making multiplanetary systems commonplace in our Galaxy. We utilize hundreds of multiplanet simulations to explore the origin of such systems, focusing on their orbital architecture. The first set of simulations assumes in situ assembly of planetary embryos, while the second explores planetary migration. After applying observational biases to the simulations, we compare them to 250+ observed multiplanetary systems, including 13 systems with planets in the habitable zone. For all of the systems, we calculate two of the so-called statistical measures: the mass concentration ( $S_c$ ) and orbital spacing ( $S_s$ ). After analytic and empirical analyses, we find that the measures are related to first order with a power law:  $S_c \sim S_s^\beta$ . The in situ systems exhibit steeper power-law relations relative to the migration systems. We show that different formation scenarios cover different regions in the  $S_s$ – $S_c$  diagram with some overlap. Furthermore, we discover that observed systems with  $S_s < 30$  are likely dominated by the migration scenario, while those with  $S_s \geq 30$  are likely dominated by the in situ scenario. We apply these criteria to determine that a majority (62%) of observed multiplanetary systems formed via migration, whereas most systems with currently observed habitable planets formed via in situ assembly. This work provides methods of leveraging the statistical measures ( $S_s$  and  $S_c$ ) to disentangle the formation history of observed multiplanetary systems based on their present-day architectures.

*Unified Astronomy Thesaurus concepts:* Exoplanet dynamics (490); Planet formation (1241); Planetary migration (2206); Exoplanet evolution (491); Protoplanetary disks (1300)

## 1. Introduction

How do planetary systems form? This is a fundamental question in astrophysics. Until recently, the solar system was the only known sample, and hence the canonical view of planet formation was developed based on this specific system (e.g., Hayashi 1981; Weidenschilling 1984).

The discovery of exoplanets entirely changes the landscape of planet formation. In particular, NASA’s Kepler mission (Borucki et al. 2010) generated a surge in observed exoplanets. We now recognize that, statistically, almost every star hosts a planet, and half of all Sun-like stars host a rocky planet in their habitable zone (e.g., Hsu et al. 2019; Bryson et al. 2021). In addition, exoplanets exhibit a wide range of architectures that contrast with the configuration of our solar system (e.g., see Winn & Fabrycky 2015; Zhu & Dong 2021, for reviews). These astonishing discoveries challenge the canonical view of planet formation.

One of the most exciting discoveries is an abundant population of a new class of exoplanets—close-in small-sized planets. Currently, at least two formation scenarios are actively investigated in the literature: the in situ and migration scenarios. The in situ scenario proposes that planets formed near where they are observed today (e.g., Ida & Lin 2010; Hansen & Murray 2012; Chiang & Laughlin 2013). This scenario is characterized by large, protoplanetary embryos that undergo at least one giant impact. Importantly, in situ formation is unlikely to explain the formation of all

multiplanetary systems (Raymond & Cossou 2014; He & Ford 2022; He & Weiss 2023).

The competing formation model is the migration scenario. In this scenario, once (proto)planets have become massive enough in the protoplanetary disk, they excite spiral density waves that provide a gravitational torque, leading to migration of the planets (Lin & Papaloizou 1979; Goldreich & Tremaine 1980; Ward 1997). Planets may lose angular momentum until they spiral into the inner edge of the disk, where the positive corotation torque halts them (e.g., Masset et al. 2006; Kley & Nelson 2012). One robust prediction of the migration scenario is that planetary systems are captured predominantly by mean motion resonances (MMRs; e.g., Terquem & Papaloizou 2007; Hellary & Nelson 2012; Izidoro et al. 2017; Ogihara et al. 2018). The period ratio distribution of observed exoplanetary systems, however, does not support this prediction (e.g., Batygin 2015; Winn & Fabrycky 2015), and hence the migration scenario alone cannot reproduce the Kepler observations either. Therefore, previous studies imply that simultaneous examination of various formation scenarios would be critical to fully reproduce the observed population of exoplanets.

Here we show that combining both the in situ and migration scenarios can lead to a better understanding of the origin of observed multiplanet systems. We provide analytic predictions to contrast characteristics of the competing formation mechanisms (Section 2). We also outline methods of determining the dominant formation mechanism of an observed planetary system based on its present-day architecture, including habitable planetary systems (Section 3). This work thus demonstrates that a huge diversity of multiplanet systems can be governed by two dominant formation scenarios: in situ and migration.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## 2. Analytical Prediction

We first provide our analytical predictions, wherein two competing formation scenarios (in situ vs. migration) are compared. To proceed, we adopt the so-called statistical measures (SMs)—quantities that characterize different aspects of a multiplanetary system (Chambers 2001). Here we discuss how the two formation scenarios can be differentiated by SMs.

### 2.1. Characteristic Separations

Different formation scenarios may end up with different separations between nearby planets. Two characteristic separations can be defined in this work:  $\Delta_{\text{in-situ}}$  and  $\Delta_{\text{mig}}$ . The former involves the in situ scenario, while the latter focuses on the planetary migration scenario.

When multiplanet systems are formed by the in situ scenario,  $\Delta_{\text{in-situ}}$  should be determined by the result of the pure gravitational interaction between neighboring protoplanets. This separation is achieved before they undergo a giant impact, that is, resultingly formed planetary systems should have a mutual separation of  $>\Delta_{\text{in-situ}}$ . Mathematically, this can be written as (e.g., Schlichting 2014)

$$\Delta_{\text{in-situ}} \simeq 2a_p \frac{v_{\text{esc}}}{v_{\text{Kep}}}, \quad (1)$$

where  $a_p$  is the semimajor axis of (proto)planets,  $v_{\text{Kep}}$  is the Keplerian velocity, and  $v_{\text{esc}} = \sqrt{2GM_p/R_p}$  is the escape velocity for the neighboring (proto)planets with comparable masses ( $M_p$ ) and radii ( $r_p$ ).

Thus, the characteristic separation between two planets set by the in situ scenario is given as

$$\begin{aligned} \frac{\Delta_{\text{in-situ}}}{a_p} &\simeq 2 \frac{v_{\text{esc}}}{v_{\text{Kep}}} \\ &\simeq 0.34 \left( \frac{M_p}{2 M_{\oplus}} \right)^{1/2} \left( \frac{R_p}{1.5 R_{\oplus}} \right)^{-1/2} \\ &\quad \times \left( \frac{M_s}{1 M_{\odot}} \right)^{-1/2} \left( \frac{a_p}{0.3 \text{ au}} \right)^{1/2}, \end{aligned} \quad (2)$$

where  $M_s$  is the mass of the host star. This indicates that if planetary systems are formed by the in situ scenario, then the spacing should be larger than  $\sim 0.3$ .

On the other hand, if planetary migration plays an important role in forming multiplanet systems, then the characteristic separation should be computed by two competing processes (Ida & Lin 2010; Arora & Hasegawa 2021): Convergent migration, which reduces the mutual spacing between neighboring protoplanets, is compensated by the gravitational repulsion between them. The common outcome is that nearby planets are captured in MMRs. Many numerical simulations show that migrating planets are captured in the 2:1 period resonances (e.g., Izidoro et al. 2017; Yu et al. 2023). Then,  $\Delta_{\text{mig}}$  can be computed using Kepler's third law as

$$\left( \frac{a_p + \Delta_{\text{mig}}}{a_p} \right)^{3/2} = 2. \quad (3)$$

Equivalently,

$$\frac{\Delta_{\text{mig}}}{a_p} = 2^{2/3} - 1 \simeq 0.59. \quad (4)$$

This suggests that if convergent planetary migration defines the characteristic separation, then it should be smaller than  $\sim 0.6$ .

### 2.2. The Statistical Measures

Characterizing multiplanet systems is not straightforward; as the number of planets in a system increases, so does the number of quantities needed to fully characterize the system. Chambers (2001) devised a series of SMs to compare the architecture of different multiplanetary systems.

We focus on two of these quantities as done in previous studies (e.g., Sun et al. 2017; Clement et al. 2019; Lykawka & Ito 2019; Arora & Hasegawa 2021; Mah et al. 2022). The first quantity is a measure of the radial mass concentration ( $S_c$ ) and is given by

$$S_c = \max \left( \frac{\sum_j M_{p,j}}{\sum_j M_{p,j} [\log_{10}(a/a_{p,j})]^2} \right), \quad (5)$$

where  $M_{p,j}$  ( $a_{p,j}$ ) is the mass (semimajor axis) of the  $j$ th planet from the host star and  $a$  is a variable denoting the distance from the host planet. A certain value of  $a$  is chosen such that  $S_c$  takes the maximum value. Compact planetary systems with similar masses and low multiplicities tend to have a high value of mass concentration. The value of  $S_c$  is essentially a measure of how evenly spaced the planets' masses are in the system.

The second quantity that we employ is the orbital spacing ( $S_s$ ), which is similar to the averaged mutual spacing normalized by the mutual Hill radius. One difference is that  $S_s$  is normalized by  $M_p^{1/4}$  (not  $M_p^{1/3}$ ). This is motivated by the results of  $N$ -body simulations that explore the stability of multiplanet systems (Chambers et al. 1996):

$$S_s = \frac{6}{N-1} \left( \frac{a_{p,\text{max}} - a_{p,\text{min}}}{a_{p,\text{max}} + a_{p,\text{min}}} \right) \left( \frac{3M_s}{2\bar{M}_p} \right)^{1/4}, \quad (6)$$

where  $N$  is the number of planets in the system,  $a_{p,\text{max}}$  ( $a_{p,\text{min}}$ ) is the maximum (minimum) semimajor axis, and  $\bar{M}_p (= \sum_j M_{p,j}/N)$  is the mean mass of planets in the system.

### 2.3. The Two-planet Case

We here consider the two-planet system case and examine how the SMs can be used to differentiate the two formation scenarios.

Our mathematical manipulation leads to a simplified expression of mass concentration (Appendix A), which is given as

$$S_c = \frac{(M_{p,1} + M_{p,2})^2}{M_{p,1}M_{p,2}} \frac{1}{[\log_{10}(a_{p,2}/a_{p,1})]^2}. \quad (7)$$

For orbital spacing,

$$S_s = 6 \left( \frac{a_{p,2} - a_{p,1}}{a_{p,2} + a_{p,1}} \right) \left( \frac{3M_s}{M_{p,1} + M_{p,2}} \right)^{1/4}. \quad (8)$$

Under the assumptions that  $M_{p,1} \simeq M_{p,2}$  and that  $\Delta a/a_{p,1} \ll 1$ , where  $a_{p,2} = a_{p,1} + \Delta a$ , the above two equations can be expressed to first order as

$$S_c \simeq 4 \left( \frac{\ln(10)}{\Delta a/a_{p,1}} \right)^2 \simeq 2.4 \times 10^2 \left( \frac{\Delta a/a_{p,1}}{0.3} \right)^{-2} \quad (9)$$

and

$$S_s \simeq 3 \left( \frac{3M_s}{2M_{p,1}} \right)^{1/4} \Delta a / a_{p,1} \\ \simeq 2.0 \times 10 \left( \frac{M_{p,1}}{2M_\oplus} \right)^{-1/4} \left( \frac{M_s}{1M_\odot} \right)^{1/4} \left( \frac{\Delta a / a_{p,1}}{0.3} \right). \quad (10)$$

We finally combine the above two equations and find that

$$S_c \simeq 4.7 \times 10^3 \left( \frac{M_{p,1}}{2M_\oplus} \right)^{-1/4} \left( \frac{M_s}{1M_\odot} \right)^{1/4} S_s^{-2}. \quad (11)$$

Equation (11) suggests that similar-mass planets that are relatively in tight orbits will exhibit an inverse-square relationship between the SMs. This is a crucial prediction and will be juxtaposed with the simulation results in later sections.

#### 2.4. Predictions

We now apply the above-simplified equations (i.e., Equations (9) and (10)) to the in situ and migration scenarios to examine whether these two scenarios can be differentiated.

As discussed in Section 2.1, if planetary systems are formed by the in situ scenario, then  $\Delta_{\text{in-situ}}/a_p \gtrsim 0.3$  (Equation (2)). This leads to  $S_c \lesssim 3 \times 10^2$  and  $S_s \gtrsim 20$ .

On the other hand, if planetary systems are formed by the migration scenario, then  $\Delta_{\text{mig}}/a_p \lesssim 0.6$  (Equation (4)). This leads to that  $S_c \gtrsim 60$  and  $S_s \lesssim 40$ .

Thus, these calculations suggest that dominant formation mechanisms could be identified by computing mass concentration and orbital spacing, and switching a formation mechanism could occur at the ranges of  $60 \lesssim S_c \lesssim 300$  and  $20 \lesssim S_s \lesssim 40$ .

These predicted ranges will be compared to the results of our simulations in Section 3.4. Despite being derived from the only two-planet case, the inequalities are robust and can be generalized to systems with three and four planets as well.

### 3. Determination of a Formation Scenario

We test the above predictions, using both simulation results and observational data available in the literature. Since exoplanetary systems observed by Kepler suffer from various observational biases, one cannot compare simulation results with observational data directly. We therefore apply observational biases to simulated planetary systems to resolve this issue. We use two kinds of simulations that study the in situ and migration scenarios. We compare simulation results with observational data to identify a formation scenario of observed systems.

#### 3.1. Observed Systems

We first describe observed systems. We gather a sample of multiplanetary systems from the catalog of the Kepler Space Telescope, following the filtering methods described in Arora & Hasegawa (2021). This observed sample is taken from the NASA Exoplanet Archive (2019)<sup>4</sup> and contains systems with at least two planets orbiting around a main-sequence star. After

the filtering, we are left with 234 observed systems (837 planets total) with two or more planets in each system.

#### 3.2. Simulation Data

We use the data from two sets of simulations. The first set contains 100 realizations of “in situ” simulations from Hansen & Murray (2013, 2015). These models assume that the protoplanetary disk has a surface density profile of  $\Sigma \sim a^{-1.5}$ , where  $a$  is the distance from the  $1M_\odot$  host star. The surface profile of the disk is then normalized to contain  $20M_\oplus$  of rocky material interior to  $a = 1$  au. The simulations are run for 10 Myr in 12 hr time steps. The in situ model assumes in-place planetary embryos and giant impact(s) to be dominant in forming multiplanetary systems.

The second set contains 538 planetary systems that undergo two-planet migration (T. Y. M. Yu et al. 2024, in preparation). These “migration” simulations explore the effects of two-planet migration on the final architecture of systems around a  $1M_\odot$  host star. The disk structure and simulation procedure are outlined in Yu et al. (2023). These simulations assume that the planets initially take on integer values between 1 and  $10M_\oplus$  and initial semimajor axes between 0.3 and 1.2 au. Initial eccentricities and inclinations are both set to 0.1, and the initial period ratio is chosen to be 2.6 between the two planets to avoid artificial capture in an MMR. The initial orbital configuration is not crucial, as it is washed out by migration.

For more details on the in situ and migration simulations, refer to Hansen & Murray (2013) and Yu et al. (2023), respectively.

We note that these simulations did not provide planet radius, so we employ the probabilistic algorithm from Chen & Kipping (2017) to calculate the radius.

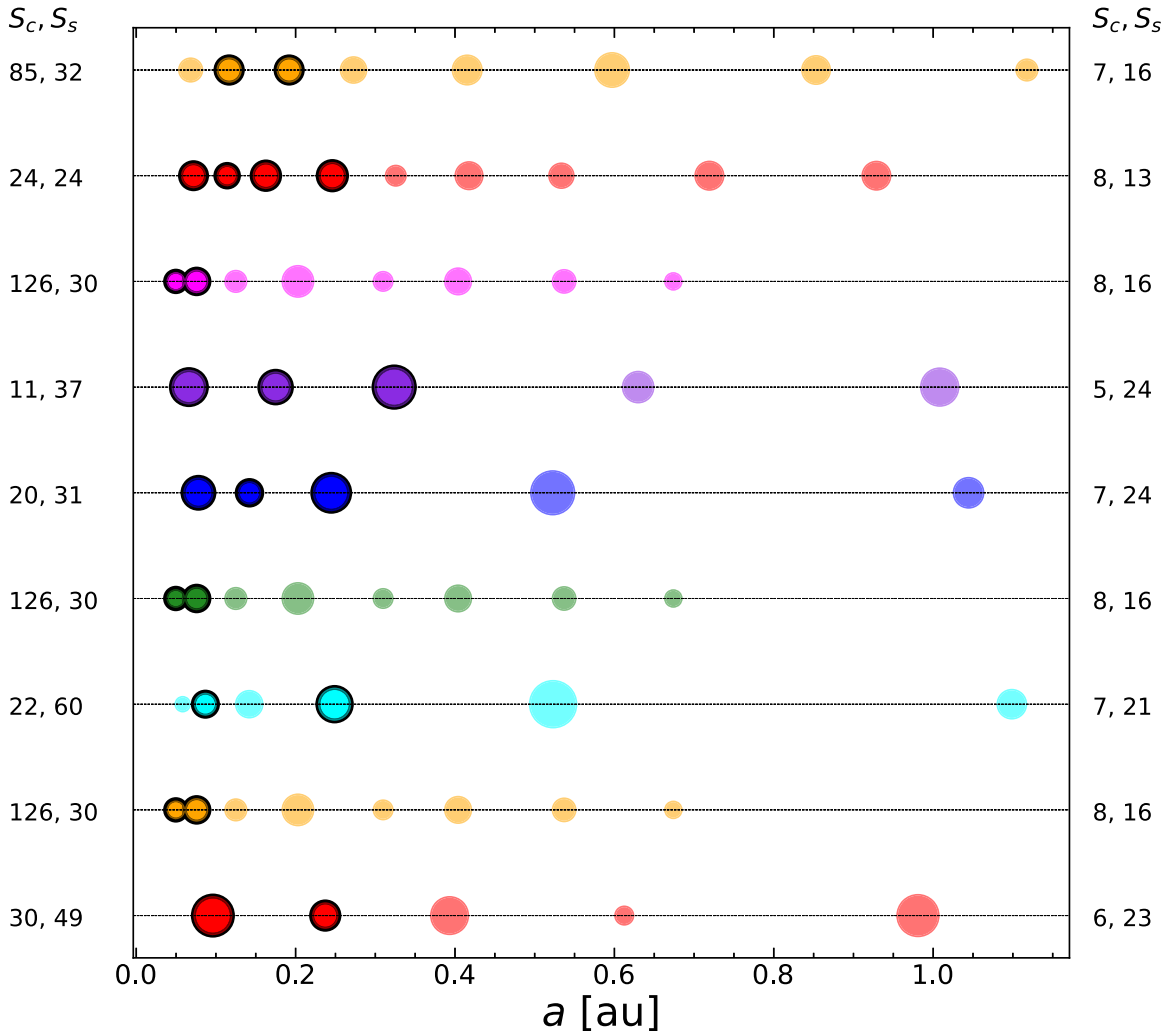
#### 3.3. Biasing Simulation Data

To meaningfully compare the simulation results to observations from Kepler, we bias the simulation data. We employ the Exoplanet Population Observation Simulator (EPOS; Mulders et al. 2018, 2019) to produce simulated observations of the theoretical systems. EPOS generates simulated Kepler observations of our multiplanetary systems, using the Kepler DR25 final catalog (Coughlin et al. 2017). The software utilizes planetary parameters (masses, semimajor axes, eccentricities, inclinations) as inputs and adopts a Monte Carlo approach to sample the geometry of the systems. From these simulated detections, EPOS supplies the various combinations of planets that could be recovered by Kepler, with the respective probabilities for each system of planets.

We plot the architecture of a subset of the in situ simulated systems before and after being biased using EPOS in Figure 1. Each row represents a multiplanetary system with the semimajor axis on the horizontal axis, and the size of each planet is scaled by its mass. The outlined planets are the ones that were detected from our simulated Kepler observations using EPOS (i.e., the “biased systems”). As expected, the more massive (i.e., larger radii) and short-period planets are recovered from the simulated observations.

We also compute the values of the SMs ( $S_c$ ,  $S_s$ ) for both biased and inherent systems and label them on the left and right sides of Figure 1, respectively. The SMs of a given system often change significantly after biasing. In most cases, both the mass concentration ( $S_c$ ) and orbital spacing ( $S_s$ ) increase,

<sup>4</sup> The source data were obtained from <https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cumulative> on 2021 January 21 at 16:17 with the size of  $9564 \times 49$  columns.



**Figure 1.** The planetary architectures of the simulated in situ multiplanetary systems before and after biasing. The planets outlined in black represent the observed planets after the simulations were biased using EPOS. All planets in a given row represent the architecture of the inherent system as a result of the in situ simulations. The sizes of the planets are scaled by the mass, and the horizontal axis represents their distance from the host star in astronomical units. We write the mass concentration ( $S_c$ ) and orbital spacing ( $S_s$ ) of the system after it was biased for each system on the left and those of the inherent system on the right. Note that the SMs change, often significantly, after observational bases are applied to the simulations.

leaving little (or no apparent) trace of the hidden, more distant planets. Particularly, biased planetary systems tend to have more concentrated mass and wider spacing than what is truly present. Additional discussions about how EPOS biases our simulated systems are outlined in Appendix B.

We find that after biasing there are 536 migration systems and 88 in situ systems with  $\geq 2$  planets that are detectable by Kepler. For the in situ samples, biasing the data greatly decreased the multiplicity of the systems. This occurs because missed planets are less massive with nonzero inclination and/or long orbital periods. On the other hand, the migration simulations originally consisted of systems with two massive, short-period planets, and hence the biasing from EPOS only removed a few systems.

### 3.4. Testing Analytical Predictions

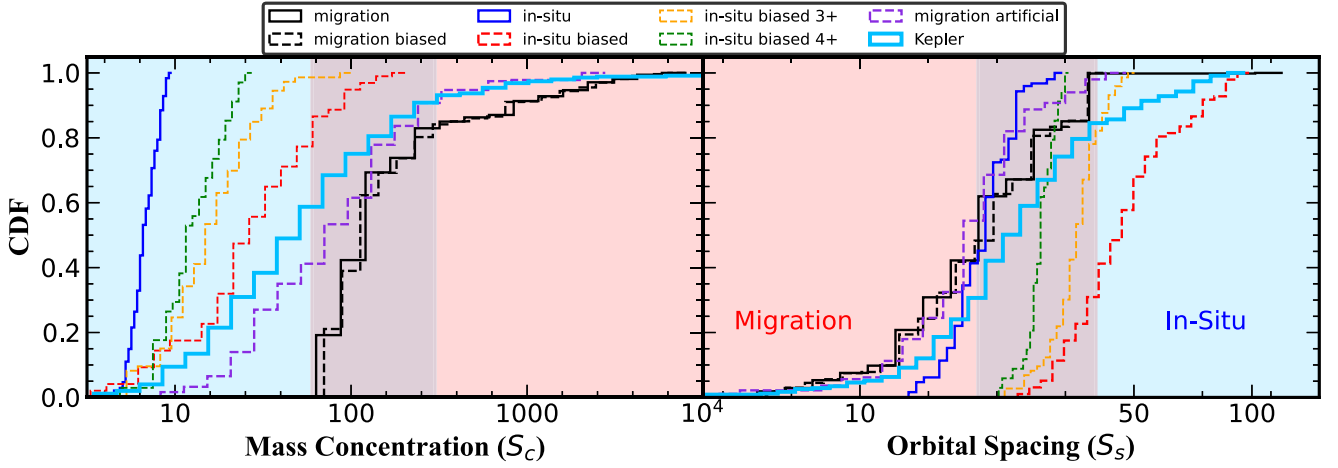
We test our analytical predictions against simulation data.

In Figure 2, we display the cumulative distribution of the SMs for the simulation data before and after biasing. For the in situ simulations, we explicitly denote biased systems that have  $\geq 3$  and  $\geq 4$  planets by the orange and green dashed

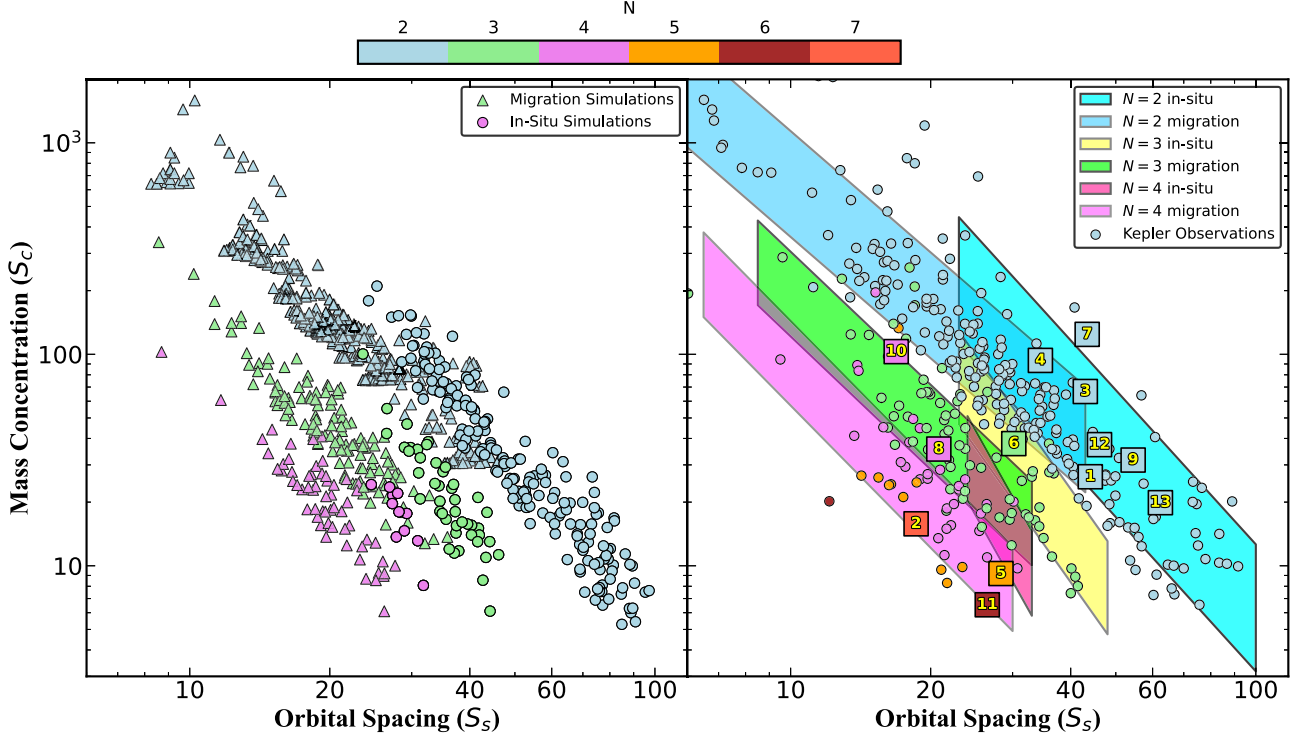
curves, respectively. The red curve displays the full set of “in situ biased” systems, which has no restrictions on the multiplicity of the biased systems.

Because the migration simulations cover only the two-planet case, we also generate an artificial set of migration systems with higher multiplicities. To pursue this, we sample the multiplicities of the artificial systems to match the multiplicity distribution of the biased in situ samples, which represent the observed systems well. Then, for a chosen multiplicity, we assign the semimajor axis of the innermost planet from the distribution of the simulated migration systems. The semimajor axis of the  $j$ th planet is computed recursively, starting with the innermost planet. We sample a semimajor axis ratio value from the distribution of the migration simulations and multiply it by the semimajor axis of the  $(j-1)$ th planet to generate the semimajor axis of the  $j$ th planet, mimicking the capture of an MMR.

The masses for each planet are chosen to be uniform in the range  $3-10 M_{\oplus}$ , similar to Yu et al. (2023). We combine 500 of these artificial samples with the two-planet migration simulations to get a combined theoretical sample, which we plot with



**Figure 2.** The CDFs of the Chambers (2001) SMs for “in situ” and “migration” simulations, before and after biasing the data. The raw migration data (black) and raw in situ data (blue) are compared to the biased migration data (dashed black) and the biased in situ data (dashed red). For the biased data, we require that the number of planets observed in the system is at least 2, making it a multiplanetary system. For the in situ data, we split it by the minimum number of observed planets, namely, three and four planets, which are shown in orange and green, respectively. The red (blue) shaded regions correspond to the ranges derived from analytic methods in Section 2.3 and are consistent with the migration (in situ) formation scenario. The artificial migration data are plotted by the dashed purple line, which shows some deviation from our analytical prediction at low  $S_c$ , while such deviation is not seen for  $S_s$ . For comparison purposes, the SMs of the Kepler data are also plotted (thick light-blue line).



**Figure 3.** The mass concentration as a function of the orbital spacing for the simulated in situ (circle) and migration (triangle) systems (left) as compared to observed Kepler systems (right). The color bar represents the multiplicity of the system. On the left, biased simulation data show a power-law relationship between  $S_c$  and  $S_s$ , and its slope is likely to be a function of multiplicity, especially for the in situ systems. On the right, the distributions of biased simulation data are denoted by the shaded regions. The Kepler observed systems are shown as the circles. The numbered squares display the SMs for 13 observed multiplanetary systems that host at least one (or more) planet(s) inside the habitable zone. They are (1) GJ 667C, (2) Trappist-1, (3) Teegarden’s Star, (4) GJ 1002, (5) Kepler-186, (6) GJ 1061, (7) Proxima Centauri, (8) K2-72, (9) GJ 273, (10) TOI-700, (11) Kepler-62, (12) LP 890-9, and (13) LHS 1140. See Appendix E for further details on these habitable planetary systems.

the dashed purple curve in Figure 2. After combining the two-planet migration simulations with the artificial migration sample, we find that the multiplicity distribution is manifestly similar to that of the biased in situ systems.

In Figure 2, we also shade the ranges of  $S_c$  and  $S_s$  as predicted from our analytical consideration (Section 2.4) and find that they are consistent with the various simulation results,

with a small region of overlap in the middle of the distribution. We confirm that the migration simulations of the two-planet case generally captured the trend of the higher-multiplicity systems. We further discuss and constrain the ranges of  $S_s$  and  $S_c$  for each formation mechanism in Section 3.5.

We finally test our predictions by examining the relationship between  $S_c$  and  $S_s$ . In Figure 3, we plot  $S_c$  as a function of  $S_s$  for

all of our biased simulation data (left panel). We color the scatter points based on their multiplicities and discover many interesting features. First, we find that, for both simulations, the populations with the same multiplicity are linear in a log–log scale. This suggests a power-law relationship between the SMs:

$$S_s \sim (S_c)^\beta. \quad (12)$$

Second, we determine the best-fit value of  $\beta$  for each of the subpopulations by performing a least-squares fit. For the simulated migration systems (squares in the left panel of Figure 3), we find that the best-fit power-law slopes are  $\beta = -2.03$ ,  $-2.20$ , and  $-2.51$  for systems with multiplicities of  $N = 2$ , 3, and 4, respectively. For the simulated in situ systems (circles in the left panel of Figure 3), we find that the best-fit power-law slopes are  $\beta = -2.42$ ,  $-3.11$ , and  $-3.93$  for systems with multiplicities of  $N = 2$ , 3, and 4, respectively. Notably, the  $N = 2$  migration systems exhibit  $\beta \approx 2$ , as predicted in Equation (11). A similar slope is observed in the  $N = 3$  migration systems, which suggests that the close-spacing assumption is a key characteristic of migration systems. Finally, we find that, within a multiplicity category, the in situ systems exhibit steeper power-law slopes than the migration systems.

In the right panel of Figure 3, we define the ranges of the parameter space that are occupied by the various simulation populations. The ranges are chosen by plotting a minimum and maximum power-law profile, based on the  $\beta$  values above, for a given multiplicity. These lines are then set to cover the minimum and maximum SM ranges reached by the simulations (scatter points in left panel of Figure 3). After this procedure, we can interpret the ranges of  $S_s$ – $S_c$  space that are occupied by the migration and in situ simulations at multiplicities 2, 3, and 4.

In the following section, we leverage the ranges to specify the dominant formation mechanism for each observed system.

### 3.5. Identifying a Formation Mechanism

It is of primary importance that the Kepler distribution is somewhere in the middle of the biased in situ and migration systems for both the  $S_c$  and  $S_s$  curves (Figure 2). This suggests that many multiplanetary systems in the observed sample can be explained by both formation histories. We aim to split the Kepler sample into two parts and identify the in situ –dominated and migration-dominated systems, based on mass concentration and orbital spacing cutoffs.

In doing so, we are guided by our theoretical predictions outlined in Section 2.4. We will show below that the true cutoff values that split the Kepler systems lie within the predicted ranges. In the following, we focus on  $S_s$  and show that it is a robust signature of the formation mechanism. We also consider a cutoff based on  $S_c$  in Appendix C.1. Additional analyses, including machine-learning classification, are summarized in Appendix C.

We test the impact of splitting the observed systems based on an  $S_s$  criterion. To find the best orbital spacing cutoff value ( $S_{s,\text{cutoff}}$ ), we consider values ranging from  $S_{s,\text{cutoff}} = 10$  to 40, motivated by the distribution in Figure 2, and find which values provide the greatest consistency between the SM cumulative distribution functions (CDFs).

We quantitatively determine  $S_{s,\text{cutoff}}$  by selecting the cutoff value that maximizes the sum of the Kolmogorov–Smirnov (KS) test  $p$ -values, which would suggest that the observed and simulated distributions are unlikely to come from different parent distributions. Here we split the Kepler sample into

systems with  $S_s < S_{s,\text{cutoff}}$  and those with  $S_s > S_{s,\text{cutoff}}$ . Then, we calculate the KS  $p$ -value between (1) the lower Kepler sample ( $S_s < S_{s,\text{cutoff}}$ ) and the biased migration systems, and (2) the higher Kepler sample ( $S_s > S_{s,\text{cutoff}}$ ) and the biased in situ systems. After sampling, we find that the best cutoff value is  $S_{s,\text{cutoff}} = 30$ , which produces  $p$ -values that are all greater than 0.05 (i.e., 95% confidence). Namely, this criterion suggests that systems with  $S_s < 30$  ( $S_s > 30$ ) are likely to have a formation history dominated by migration (in situ), which we display in Figure 4. Leveraging this criterion, we find that 62% of our observed systems likely exhibited planetary migration as their dominant formation mechanism.

A similar but more restricted argument can be made using  $S_c$  (see Appendix C.1). By splitting the Kepler sample based on these cutoff values, we discover that it is also consistent with the simulated period ratio distribution, as well as other orbital characteristics, which we outline in Appendix D.

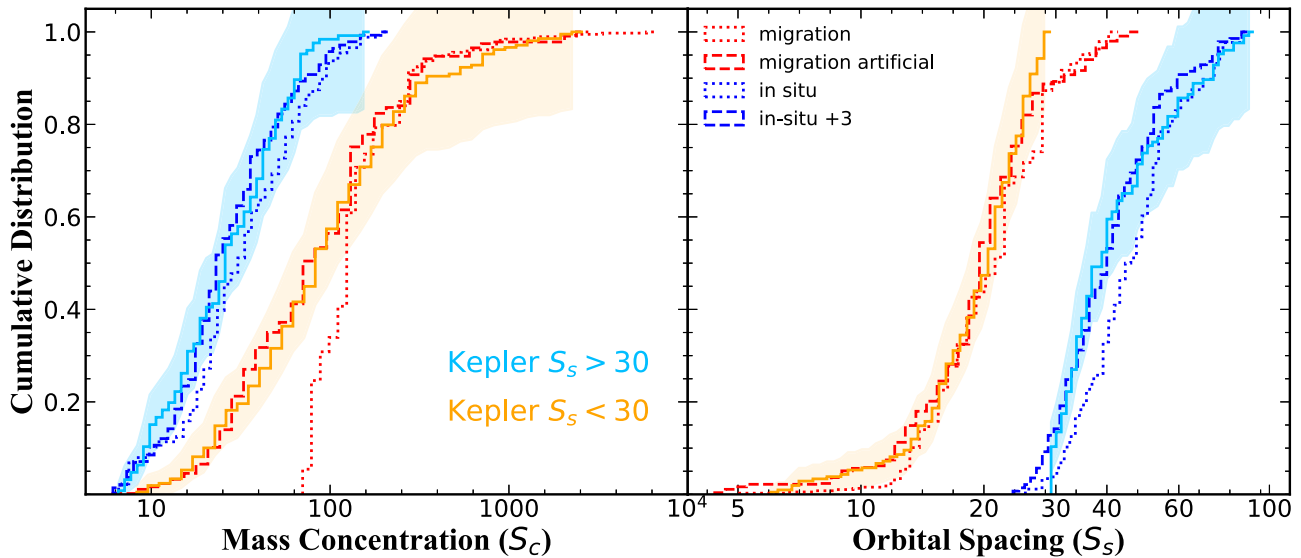
By combining the empirical SM ranges above with the theoretical power-law ranges provided in Figure 3, we can predict the dominant formation mechanism of nearly all multiplanetary systems in our sample. Leveraging the SMs in this way has shown to be highly robust in gaining insights into the formation mechanism of any observed multiplanetary systems.

One can infer the dominant formation mechanism using the following systematic process:

1. If the calculated ( $S_s$ ,  $S_c$ ) values lie in a region of the  $S_s$ – $S_c$  diagram (Figure 3) that is unique to one formation mechanism (i.e., in a shaded region without overlap) in accord with the multiplicity of the system, then that is the dominant formation mechanism.
2. If ( $S_s$ ,  $S_c$ ) values lie in a region of the  $S_s$ – $S_c$  diagram that is occupied by either multiple shaded regions with the same multiplicity or unshaded regions, then employ the  $S_s = 30$  criteria. Namely, if  $S_s < 30$ , then the formation of the systems was likely dominated by planetary migration. Conversely, if  $S_s > 30$ , the formation of the systems was likely dominated by in situ assembly.

Identifying a formation mechanism of planetary systems with high multiplicity only by  $S_s$  and/or  $S_c$  could be hard. In the current analysis, such identification becomes possible largely for planetary systems that exhibit two to three planets after biasing (Figure 3). As described in Section 3.3, biasing affects planetary systems formed by in situ considerably, while it does not for systems formed by migration. This occurs because the former systems tend to have planets that are less massive (i.e., smaller) with nonzero inclinations and/or long orbital periods. These planets are susceptible to observational biases and have a high chance of not being detected by transit observations. The resulting change in orbital architecture is still significant enough to trace the formation history, as shown in this work.

For high-multiplicity planetary systems, however, if most of the planets in the systems are observed, both the formation mechanisms lead to tightly packed systems; observed planets should be reasonably massive, and their inclinations/semimajor axes should be small. This configuration makes it difficult to specify the formation mechanism only by  $S_s$  and/or  $S_c$ . In fact, for systems with  $N = 4$ , the in situ simulations largely overlap with other regions (Figure 3), therefore introducing ambiguity into that region of the  $S_s$ – $S_c$  plot. Furthermore, most of such high-multiplicity systems have  $S_s < 30$ , which make the second



**Figure 4.** The cumulative distributions of the mass concentration ( $S_c$ ; left column) and orbital spacing ( $S_s$ ; right column) of the biased simulation data compared to the observed Kepler sample, which are both split into two subgroups. The dotted blue lines are the biased in situ simulated systems, and the dashed blue lines are in situ systems selected for more systems with multiplicities greater than 3. The dotted red lines are the distributions of the biased migration simulation data, and the dashed red lines are the artificial migration systems generated to include higher multiplicities (see Section 3.4 for details). We compare this to the observed Kepler systems with different cutoffs. We split the observed systems based on those with  $S_s < 30$  (orange) and those with  $S_s > 30$  (light blue). For the observed distributions, we include shaded regions that represent Poisson error bars for each bin.

criterion less effective for such systems. In addition, we find that most (97%) of the in situ systems without biases would be classified as having a formation mechanism consistent with migration, only based on their  $S_s$  value. Different quantities (e.g., bulk and atmospheric compositions) are clearly needed to reliably identify a formation mechanism of planetary systems with high multiplicity. The current transiting technologies do not have the capability to observe small-sized planets with high inclination and/or large orbital periods. Therefore, our focus on the biased architecture provides a strong prediction to uncover the formation history of planetary systems observed currently.

### 3.6. Application to Planetary Systems in Habitable Zones

In the previous subsection, we established methods to gain insight into planetary formation history. Here we apply these criteria to multiplanetary systems that have at least one planet in the habitable zone (hereafter “habitable systems”) to determine their dominant formation mechanisms. The habitability criteria we employ are conservative and seek planets that are likely to exhibit a rocky composition and support surface liquid water (habitable zones determined by Kopparapu 2013; Kopparapu et al. 2013, for M dwarfs and other main-sequence stars, respectively). For our systems,<sup>5</sup> we look for planets with  $0.5 < R_p/R_\oplus \leq 1.6$  or  $0.1 < M_{p,\min}/M_\oplus \leq 3$ , or slightly more optimistically,  $1.6 < R_p/R_\oplus \leq 2.5$  or  $3 < M_{p,\min}/M_\oplus \leq 10$ . We further detail the identified systems and information about their habitability in Appendix E.

We calculate the SMs ( $S_s$ ,  $S_c$ ) of 13 habitable systems identified from the above criteria. For these planets, if there was no estimate provided for the planet mass ( $M_p$ ), we determine the mass in two ways, depending on the method of detection. If the planet was detected using radial velocity methods and has a reported  $M_p \sin i$  value, we average this over an inclination

distribution that is uniform in  $\cos i$  to arrive at a median  $M_p$  estimate. On average, this leads to a mass estimate that is 16% larger than the reported  $M_p \sin i$  (e.g., Zechmeister et al. 2019). If the planet was detected using any other method, we apply the Chen & Kipping (2017) mass–radius relation to estimate the  $M_p$  for planets that only have a reported radius. This follows naturally because the Chen & Kipping (2017) mass–radius relation was employed to derive radius estimates for the simulations. We often apply this method for planets detected by transit. The SMs of the 13 systems, along with their predicted formation mechanisms, are listed in Table 1.

Through the criteria discussed in the previous section, we predict that systems GJ 667C (1), GJ 1061 (6), GJ 273 (9), LP 890-9 (12), and LHS 1140 (13) were formed via in situ assembly. From the second criterion, we infer that systems K2-72 (8) and TOI-700 (10), which exhibit  $S_s < 30$ , were dominated by planetary migration.

We note that if the multiplicity is not covered by the simulated sample, as is the case for systems Trappist 1 (2), Kepler-186 (5), and Kepler-62 (11), then we cannot make a strong prediction regarding the formation because such systems are not represented in our biased simulation data.

Proxima Centauri (7) lies in a region of Figure 3 that is unoccupied by our simulations, so we leave the formation mechanism ambiguous. Teegarden’s Star (3) and GJ 1002 (4) lie in regions of Figure 3 that are occupied by both migration and in situ systems, so we cannot make any definite predictions about their formation, though we suggest that the former may have experienced early migration followed by in situ assembly, as described in Hansen & Murray (2012). Moreover, modern telescope technology strongly favors the detection of close-in planets and is not currently powerful enough to identify Earth-like habitable planets out to 1 au. Therefore, the applications of our method to determine the formation mechanism of habitable planets are restricted to observable planets that are relatively close-in, similar to the population described above.

<sup>5</sup> For the search, we used the Habitable Worlds Catalog: <https://phl.upr.edu/hwc>.

**Table 1**  
Statistical Measures of Habitable Systems

Planetary System	$S_s$	$S_c$	Predicted Formation	Multiplicity
1. GJ 667C	44.1	26.4	in situ	2
2. Trappist-1	18.6	15.8	...	7
3. Teegarden's Star	43.0	66.7	...	2
4. GJ 1002	34.4	93.6	in situ/migration	2
5. Kepler-186	28.3	9.2	...	5
6. GJ 1061	30.2	37.8	in situ	3
7. Proxima Centauri	43.4	124.2	...	2
8. K2-72	20.8	35.6	migration	4
9. GJ 273	54.5	31.6	in situ	2
10. TOI-700	16.9	103.0	migration	4
11. Kepler-62	26.5	6.5	...	5
12. LP 890-9	46.1	37.4	in situ	2
13. LHS 1140	62.4	19.9	in situ	2

#### 4. Conclusions and Summary

In this study, we analyze two sets of planet formation simulations to identify unique characteristics within each one. The first set assesses planetary migration, while the other tests in situ assembly as the dominant formation mechanism. We introduce biases to model artificial detections akin to observations from the Kepler Space Telescope. For each biased system, we calculate the SMs, which include mass concentration ( $S_c$ ) and orbital spacing ( $S_s$ ). After biasing the simulations, we compare them to a catalog of over 250 Kepler systems, aiming to predict the dominant formation mechanism of these observed systems, with a particular focus on systems with habitable planets. We summarize our analysis and findings as follows:

1. Through theoretical calculations, we predict that if a system has a formation history dominated by planetary migration, it will exhibit  $S_c \gtrsim 60$  and  $S_s \lesssim 40$ , and if it was dominated by in situ assembly, it will exhibit  $S_c \lesssim 300$  and  $S_s \gtrsim 20$ . We further expand on the two-planet case to derive a relationship between  $S_s$  and  $S_c$ , finding that, to first order,  $S_c \sim S_s^{-2}$ , assuming that the planets are in a tight orbit and have similar masses.
2. We calculate and compare the SMs for both the intrinsic and biased systems and find that they can differ significantly. Both the  $S_c$  and  $S_s$  generally increase after biasing. The architecture before and after is displayed in Figure 1, and the SMs before and after are displayed in Figure 2.
3. We create a scatter plot of the SMs of simulated systems and find that they obey power-law relations,  $S_c \sim S_s^\beta$  (left panel of Figure 3). By performing a least-squares fit, we find that  $\beta = -2.03, -2.20,$  and  $-2.51$  for migration systems with multiplicities of  $N=2, 3,$  and  $4,$  respectively. For the simulated in situ simulations, we find that  $\beta = -2.42, -3.11,$  and  $-3.93$  for systems with multiplicities of  $N=2, 3,$  and  $4,$  respectively. Notably, the  $N=2$  migration systems exhibit  $\beta \approx 2$ , which is consistent with the assumptions and analytic predictions provided earlier.
4. We attempt to split our entire Kepler sample of multiplanetary systems by a criterion that depends on either  $S_c$  or  $S_s$ . We find that separating the Kepler sample at about  $S_s = 30$  is optimal to discriminate between formation mechanisms. From Figure 4, we conclude that systems

with  $S_s < 30$  ( $S_s > 30$ ) are likely to have a formation history dominated by migration (in situ).

5. Finally, we leverage our analysis of the SMs to predict the dominant formation mechanism of 257 observed multiplanetary systems, including 13 with one or more planets in the habitable zone. Based on the  $S_s$  criteria, we conclude that 62% of our observed sample experienced planetary migration. Based on Figure 3 and Table 1, we also find that most of our habitable systems formed via in situ assembly. This could suggest that the dynamical history that comes with in situ assembly might increase the formation and detection rates of terrestrial planets in the habitable zone.

In this work, we demonstrate the profound capabilities of the SMs to predict the formation history of multiplanetary systems. We also find that the SMs of the biased systems differ from those of the intrinsic system (e.g., Figure 1). If these changes are moderately predictable, one can disentangle the characteristics of the intrinsic systems based on their observed architecture. We do not pursue these methods of ‘‘debiasing’’ in this work but note that it could be possible for future work.

#### Acknowledgments

We thank the anonymous referee for a constructive report. C. S. thanks the SURF@JPL program. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). Y.H. is supported by JPL/Caltech.

#### Appendix A

##### A Simplified Expression of Mass Concentration for the Two-planet Case

We here briefly describe how a simplified expression of mass concentration can be derived for the two-planet case.

Mass concentration for the two-planet case is written as (Equation (5))

$$S_c = \max \left( \frac{M_{p,1} + M_{p,2}}{M_{p,1} [\log_{10}(a/a_{p,1})]^2 + M_{p,2} [\log_{10}(a/a_{p,2})]^2} \right). \quad (\text{A1})$$

A maximum value can be achieved when the denominator takes a minimum value. Taking the derivative of the denominator, we find that it becomes possible when

$$\log_{10} a = \frac{M_{p,1} \log_{10}(a_{p,1}) + M_{p,2} \log_{10}(a_{p,2})}{M_{p,1} + M_{p,2}}. \quad (\text{A2})$$

Plugging the above  $\log_{10} a$  into Equation (A1), then mass concentration is expressed as

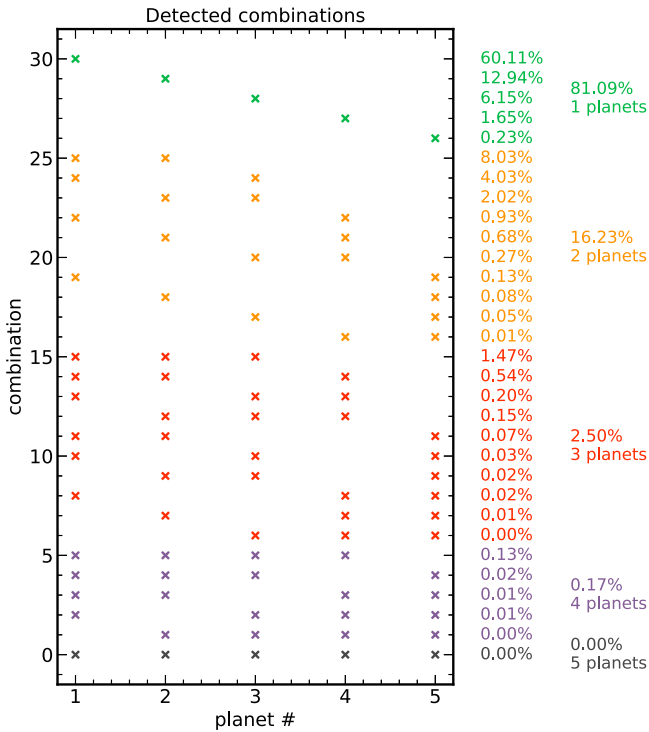
$$S_c = \frac{(M_{p,1} + M_{p,2})^2}{M_{p,1} M_{p,2}} \frac{1}{[\log_{10}(a_{p,2}/a_{p,1})]^2}. \quad (\text{A3})$$

#### Appendix B

##### Additional Discussions about the Outcome of EPOS

We provide a brief additional examination of the outcome of EPOS.





**Figure 5.** Detection combinations and probabilities for an example multiplanetary system. Observing the innermost planet is the most likely scenario (60.11%). Observing the inner two planets is the most probable (8%) multiplanet observation to be made by Kepler.

After biasing by EPOS, the multiplanetary systems are often reduced to the systems that have planets with larger masses and shorter periods.

In Figure 5, we display the probabilities of different combinations of planets being observed by Kepler. In this example, it is most likely (81.09%) that only one planet will be recovered, with the highest probability of  $\sim 60\%$  that only the closest planet will be observed. However, there is a 16.23% chance that two planets will be recovered, with the most visible combination being the inner two planets (planets 1 and 2). The probability is highly dependent on the orientation of the system (Mulders et al. 2018). We also find that the shorter-period planets, regardless of the planet’s mass, have higher probabilities of being detected. Moreover, the planets with relatively low mutual inclinations ( $i \lesssim 5^\circ$ ) are more easily observed with simulated Kepler transit detections.

## Appendix C Classification of Exoplanetary Systems

We provide additional discussions about how observed exoplanetary systems can be classified so that their formation origins are identified.

### C.1. Splitting by $S_c$

By inspection of Figure 2, we find that the minimum mass concentration of the two-planet migration systems is  $S_c \sim 70$ . Therefore, we first filter the Kepler multiplanetary systems solely under the condition that  $S_c \geq 70$ . The total of 301 planets in 139 systems satisfy this criterion, making them potentially more consistent with migration-dominated formation. For the remaining systems with  $S_c < 70$ , we further constrain them to have  $S_s$  bounds that are matched with the limitations of the

in situ simulations. The total of 309 planets in 133 systems satisfy these criteria, making them potentially more consistent with an in situ-dominated formation scenario.

In Figure 6 we plot both of the aforementioned subgroups derived from our Kepler sample alongside the biased theoretical systems. We first plot the curve of the SMs for the in situ (dashed blue) and migration (dashed red) simulations after they have been biased (see Section 3.3 for details on the biasing). Next, in light blue, we plot the Kepler systems with  $S_c < 70$  and  $S_s$  within the range of the in situ data. In orange, we show the Kepler systems with  $S_c \geq 70$ . The shaded regions for both of these curves represent the Poissonian error ( $\sqrt{N}$ ) for each bin.

We find that Kepler systems with  $S_c \geq 70$  have both  $S_c$  and  $S_s$  distributions that are broadly consistent with those from the migration simulations. Namely, we can leverage this consistency to determine that the observed multiplanetary systems in our Kepler sample with  $S_c \geq 70$  reflect a formation history dominated by planetary migration. Moreover, the remaining Kepler systems (light blue in Figure 6) have an  $S_c$  distribution consistent with the simulated in situ systems.

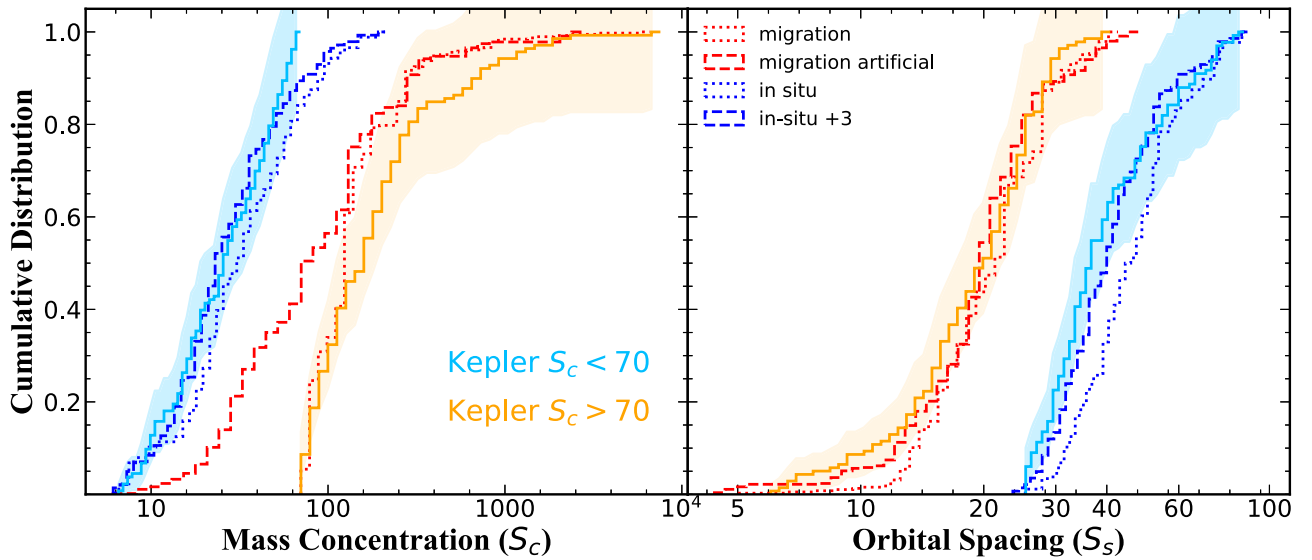
We find that the in situ systems with higher multiplicities occupy the region of abundance  $S_s \in \{25, 50\}$ . This could suggest that the Kepler systems with higher multiplicities, which are missing from the sample, would also occupy this region. We test this hypothesis to find that, when we invoke a larger number of higher-multiplicity ( $\geq 3$ ) systems into our biased in situ sample, the curves overlap nearly identically. The dashed green curves in Figure 6 represent the in situ simulations sampled so that their proportion of three-planet systems matches the ratio in our observed sample.

Although this  $S_c$  cutoff was initially determined from the minimum  $S_c$  of the migration simulations, we find that it is statistically optimized. Namely, we perform the KS test for the cumulative distributions of  $S_c$  and  $S_s$  compared to the distributions of the observed Kepler systems to determine which  $S_c$  cutoff between  $S_c \in \{50, 90\}$ . From this,  $S_c = 70$  gives the largest KS test  $p$ -values. The  $S_c = 70$  cutoff produces  $p$ -values greater than 0.05 (assuming 95% confidence), which does not support the hypothesis that the two distributions came from different parent distributions.

### C.2. Machine-learning Classification

We verify our classification in Section 3.5 and Appendix C.1, using machine learning. We test various machine-learning methods, trained on the simulation data, to be applied to the Kepler data. These supervised methods are trained on the following features of the multiplanetary system: the mass concentration ( $S_c$ ), the orbital spacing ( $S_s$ ), the average period ( $P_{av}$ ), the mass ratio ( $M_{out}/M_{in}$ ), the period ratio ( $P_{out}/P_{in}$ ), and the multiplicity. We present a corner plot of these parameters and their correlations with each other in Figure 7.

After training and testing the model with the simulation data, we find that the accuracy is effectively unchanged if we remove the last three features. With only the first three features in consideration, we employ several machine-learning algorithms and assess their capability to predict the dominant formation mechanism of multiplanetary systems.



**Figure 6.** The cumulative distributions of the mass concentration ( $S_c$ ; left column) and orbital spacing ( $S_s$ ; right column) of the biased simulation data, as well as the observed Kepler sample, which is split into two subgroups. The curves are the same as in Figure 4, but here we split the Kepler sample by those with  $S_c > 70$  (orange) and those with  $S_c < 70$  (light blue).

### C.2.1. $k$ -Nearest Neighbors

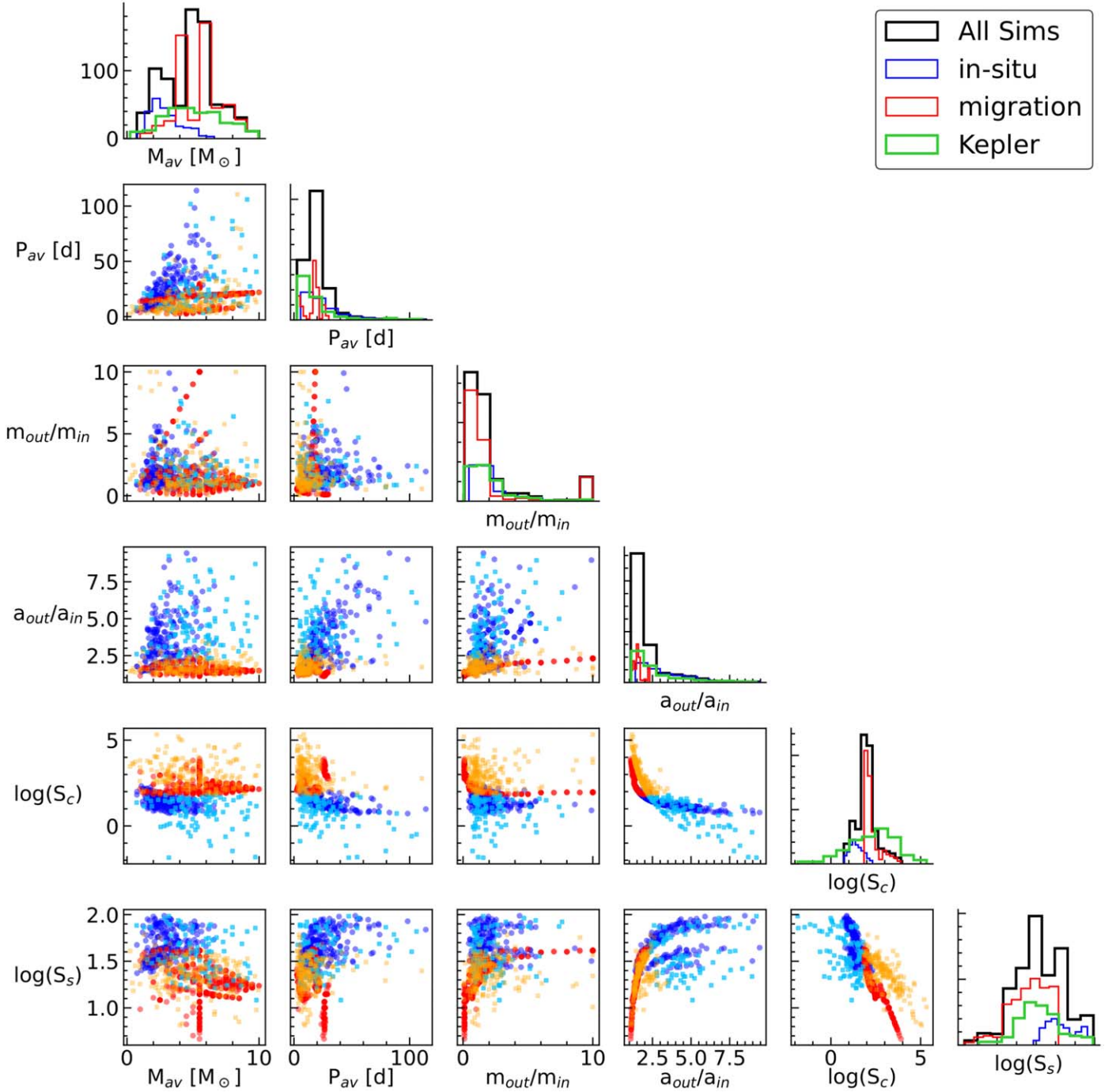
The first machine-learning model we use is the  $k$ -Nearest Neighbors (kNN) algorithm. This model as a method of classifying detections of exoplanets has been deemed highly fruitful in previous work (Herur et al. 2022). After preprocessing, training, and testing our data with the kNN model, we reach an accuracy of 96%. We discover that mass concentration is one crucial predictor of the formation mechanism of the system. However, the fact that the average period is highly correlated with the formation mechanism might be a result of the bias in the migration simulations. Namely, these simulations exhibit smaller periods by design, so we can ignore this feature as well. Interestingly, a majority ( $\sim 65\%$ ) of systems are classified as having a formation process dominated by migration of the inner planets. Most importantly, the kNN model also supports that the mass concentration and orbital spacing are the two most important features in predicting the dominant formation mechanism.

### C.2.2. Neural Network Model

As an alternative to the already highly accurate kNN model, we also test a manually designed neural network (NN) for the binary classification process. The NN is made with only one hidden layer containing 6464 nodes to avoid overfitting, especially considering our small number of features and relatively low sample size. After training the model for 20 epochs, we reach an accuracy of 92.8.

### C.2.3. Other ML Classifiers

To be complete, we also test Support Vector Machine and logistic regression models to see whether any of them are more accurate than our other algorithms. With optimization, we find that these models are unable to reproduce accuracies as high as the previous models, so we do not apply them to the Kepler data.



**Figure 7.** Triangle plot of the parameters used to train the kNN model to predict the dominant formation mechanism. The features of the multiplanetary systems that were inputted into the various ML models include the (1) average period ( $P_{av}$ ), (2) mass ratio between the farthest and closest planet ( $m_{out}/m_{in}$ ), (3) semimajor axis ratio between the farthest and closest planet ( $a_{out}/a_{in}$ ), (4) mass concentration ( $S_c$ ), and (5) orbital spacing ( $S_s$ ).

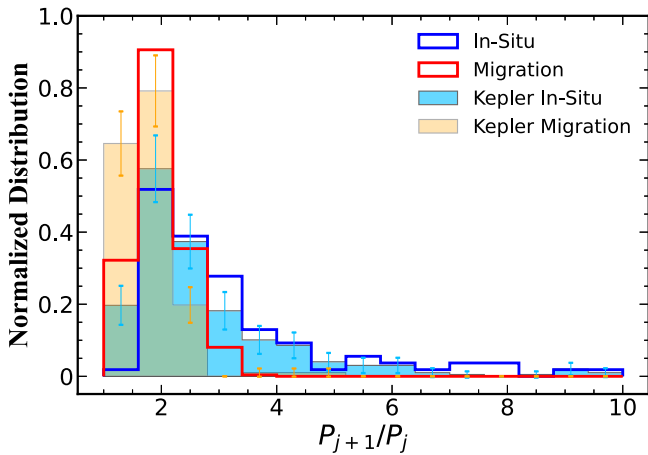
## Appendix D

### Characteristics of Different Formation Mechanisms

Planets that migrate inward are likely to exhibit long resonant chain structures where the period ratios of neighboring planets are simply integer ratios (e.g., Terquem & Papaloizou 2007; Raymond et al. 2008; Izidoro et al. 2017; Ormel et al. 2017; Kajtazi et al. 2023). We look for planets close to such MMRs within the Kepler systems. We identify a pair of adjacent planets as having a resonant structure if their period ratio ( $P_{j+1}/P_j$ ) is within 10% of the 2:1, 4:3, or 3:2 resonant

states. See Figure 8 for the distributions of period ratios for all of our models and observations.

We find that the Kepler migration systems—with radial mass concentrations larger than 70—exhibit more systems with resonant structures than the Kepler in situ systems. Similar to the migration simulations, the Kepler migration planets have peaked period ratios near the resonant values of 2/1, 4/3, and 3/2. Specifically, we find that 73% (118/162) of the Kepler migration planet pairs exhibit one of these primary MMRs, whereas only 15% (26/176) of the Kepler in situ systems have planetary pairs with these integer period ratios.



**Figure 8.** The normalized density distribution of the period ratios ( $P_{j+1}/P_j$ ,  $j \in \mathbb{Z}$ ) for all adjacent planets. The migration (solid red) and in situ (solid blue) simulations are compared to the Kepler migration systems (light orange) and Kepler in situ systems (light blue). The observed systems have Poisson error bars ( $\sqrt{n}$  for each bin) overplotted.

This feature appearing in orbital characteristics can be viewed as an additional confirmation that our classification of observed exoplanetary systems by the SMs is reasonable.


### Appendix E Habitable Systems

1. *GJ 667C* is proposed to host six (or even seven) planets with varying masses and orbital periods (Anglada-Escudé & Butler 2012; Delfosse et al. 2013). Doppler measurements and fitting analyses identify three super-Earth systems in the habitable zone, with highly constrained planetary parameters (Anglada-Escudé et al. 2013). The SMs ( $S_s$ ,  $S_c$ ) for this system are (41.9, 26.4).
2. *Trappist-1* is a system of seven temperate terrestrial planets, with the six inner planets forming a near-resonant chain (Gillon et al. 2017). From the resonant architecture, the formation is likely to be more consistent with a migration scenario, as outlined in Gillon et al. (2017). However, the migration simulations that we employ are restricted to only two-planet systems, which makes this system a poor fit to compare to our sample. Regardless, we find that the orbital spacing of the Trappist-1 system is in the range of the migration simulations and has no overlap with the in situ simulations. The SMs ( $S_s$ ,  $S_c$ ) for this system are (18.6, 15.8).
3. *Teegarden's Star* is a star in the nearby solar neighborhood with evidence from over 200 radial velocity measurements for two Earth-mass planet candidates. The CARMENES search for exoplanets identified these planets, which have orbital periods of 4.91 and 11.4 days, both with a minimum mass of  $1.1 M_\oplus$  (Zechmeister et al. 2019). One planet is in the conservative habitable zone, with both being in the optimistic habitable zone. The SMs ( $S_s$ ,  $S_c$ ) for this system are (43.0, 66.7).
4. *GJ 1002* was reported to host two temperate Earth-size planets in its habitable zone through radial velocity detections. The planet GJ 1002 b (GJ 1002 c) has an orbital period of 10 days (21 days), with minimum masses of  $1.08 M_\oplus$  ( $1.38 M_\oplus$ ) (Suárez Mascareño et al. 2023). The SMs ( $S_s$ ,  $S_c$ ) for this system are (32.7, 93.6).

5. *Kepler-186* is a main-sequence star, which initially had four planets detected with radii less than  $1.5 R_\oplus$  and orbital periods ranging from 3.9 to 22.4 (Lissauer et al. 2014; Rowe et al. 2014). A fifth planet, Kepler-186f, was then detected in the conservative habitable zone of the host star (Kopparapu et al. 2013; Tenenbaum et al. 2013; Quintana et al. 2014) with an estimated intermediate mass of  $1.44 M_\oplus$  and period of 130 days (Quintana et al. 2014). The SMs ( $S_s$ ,  $S_c$ ) for this system are (28.3, 9.2).
6. *GJ 1061* is a low-mass star with three dynamically stable planet candidates, with the potential for a fourth hidden planet, though the fourth signal may also be due to stellar rotation (Dreizler et al. 2020). The third planet, GJ 1061d, lies within the habitable zone of the star and exhibits an equilibrium temperature similar to that of Earth. The planets have periods of 3.2, 6.7, and 13.0 days and minimum masses of 1.4, 1.8, and  $1.7 M_\oplus$  (Dreizler et al. 2020). The SMs ( $S_s$ ,  $S_c$ ) for this system are (28.7, 37.8).
7. *Proxima Centauri* is the nearest known star to our solar system and has been proposed to host two or three planets (Anglada-Escudé et al. 2016; Li et al. 2017), though the presence of the third planet has recently been challenged (Artigau et al. 2022). To be conservative, we only consider the first two planets. These two planets have orbital periods of 5.1 and 11.2 days and minimum masses of 0.26 and  $1.07 M_\oplus$  (Suárez Mascareño et al. 2020; Faria et al. 2022). The second planet lies in the habitable zone of the star. The SMs ( $S_s$ ,  $S_c$ ) for this system are (41.3, 124.2).
8. *K2-72* is located 67 pc away and is detected to host four planets with sizes similar to that of Earth (Crossfield et al. 2016; Dressing et al. 2017). The third planet in this system, K2-72 c, lies in the conservative habitable zone. The SMs ( $S_s$ ,  $S_c$ ) for this system are (35.6, 19.8).
9. *GJ 273* is the second-nearest known planetary system with two detections, Earth-mass planets detected via radial velocity methods (Astudillo-Defru et al. 2017). The second of these planets, GJ 273 b, lies in the conservative habitable zone. The SMs ( $S_s$ ,  $S_c$ ) for this system are (31.6, 51.8).
10. *TOI-700* is an M2.5 dwarf star that hosts three planets discovered by TESS, with four planets (Kirkpatrick et al. 1991; Gilbert et al. 2020, 2023; Rodriguez et al. 2020). Two of these planets, TOI-700 d and TOI-700 e, are Earth-sized planets in the habitable zone of the star (Gilbert et al. 2020, 2023; Rodriguez et al. 2020). The SMs ( $S_s$ ,  $S_c$ ) for this system are (103.0, 16.0).
11. *Kepler-62* is star that hosts five Earth-sized planets (Borucki et al. 2013). Two of the planets, Kepler-62e and Kepler-62f, are super-Earths in the habitable zone that are estimated to have rocky compositions with mostly solid water (Borucki et al. 2013). Kepler-62f was confirmed to not be a true detection with updated planetary properties by more recent Gaia results (Borucki et al. 2019). The SMs ( $S_s$ ,  $S_c$ ) for this system are (6.5, 26.5).
12. *LP 890-9* was detected by TESS (Muirhead et al. 2018) and later discovered to host two temperate super-Earths via transit detection (Delrez et al. 2022). The outer planet, LP 890-9 c, lies in the conservative habitable zone. The SMs ( $S_s$ ,  $S_c$ ) for this system are (37.4, 46.1).
13. *LHS 1140* is a nearby (10.5 pc; Gaia Collaboration et al. 2018) planetary system with two known small, short-

period planets (LHS 1140 c,  $P_c \sim 3.77$  days; LHS 1140 b,  $P_b \sim 24.7$  days). The planets were detected using a combination of transit and radial velocity methods (Mayor et al. 2003; Nutzman & Charbonneau 2008). LHS 1140 b lies within the habitable zone of the star and has recently updated planet properties. Lillo-Box et al. (2020) combined the power of ESPRESSO and TESS to further constrain the masses to  $6.48 \pm 0.46 M_{\oplus}$  for LHS 1140 b and  $1.78 \pm 0.17 M_{\oplus}$  for LHS 1140 c. Moreover, fitting-analysis methods suggest another planet on a 78.9-day period and a mass of  $4.8 \pm 1.1 M_{\oplus}$ , though it has not been confirmed, so we do not include it here. We note that including this third planet does not change the inferred formation mechanism based on our assessments. The SMs ( $S_s, S_c$ ) for this system are (62.4, 19.9).

### ORCID iDs

Cheyenne Shariat  <https://orcid.org/0000-0003-1247-9349>  
 Yasuhiro Hasegawa  <https://orcid.org/0000-0002-9017-3663>  
 Bradley M. S. Hansen  <https://orcid.org/0000-0001-7840-3502>  
 Tze Yeung Mathew Yu  <https://orcid.org/0000-0002-5228-7176>  
 Renyu Hu  <https://orcid.org/0000-0003-2215-8485>

### References

- Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, *Natur*, 536, 437  
 Anglada-Escudé, G., & Butler, R. P. 2012, *ApJS*, 200, 15  
 Anglada-Escudé, G., Tuomi, M., Gerlach, E., et al. 2013, *A&A*, 556, A126  
 Arora, U., & Hasegawa, Y. 2021, *ApJL*, 915, L21  
 Artigau, É., Cadieux, C., Cook, N. J., et al. 2022, *AJ*, 164, 84  
 Astudillo-Defru, N., Forveille, T., Bonfils, X., et al. 2017, *A&A*, 602, A88  
 Batygin, K. 2015, *MNRAS*, 451, 2589  
 Borucki, W., Thompson, S. E., Agol, E., & Hedges, C. 2019, arXiv:1905.05719  
 Borucki, W. J., Agol, E., Fressin, F., et al. 2013, *Sci*, 340, 587  
 Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Sci*, 327, 977  
 Bryson, S., Kunimoto, M., Kopparapu, R. K., et al. 2021, *AJ*, 161, 36  
 Chambers, J. E. 2001, *Icar*, 152, 205  
 Chambers, J. E., Wetherill, G. W., & Boss, A. P. 1996, *Icar*, 119, 261  
 Chen, J., & Kipping, D. 2017, *ApJ*, 834, 17  
 Chiang, E., & Laughlin, G. 2013, *MNRAS*, 431, 3444  
 Clement, M. S., Kaib, N. A., Raymond, S. N., Chambers, J. E., & Walsh, K. J. 2019, *Icar*, 321, 778  
 Coughlin, J., Thompson, S. E. & Kepler Team 2017, AAS Meeting Abstracts, 230, 102.04  
 Crossfield, I. J. M., Ciardi, D. R., Petigura, E. A., et al. 2016, *ApJS*, 226, 7  
 Delfosse, X., Bonfils, X., Forveille, T., et al. 2013, *A&A*, 553, A8  
 Delrez, L., Murray, C. A., Pozuelos, F. J., et al. 2022, *A&A*, 667, A59  
 Dreizler, S., Jeffers, S. V., Rodríguez, E., et al. 2020, *MNRAS*, 493, 536  
 Dressing, C. D., Vanderburg, A., Schlieder, J. E., et al. 2017, *AJ*, 154, 207  
 Faria, J. P., Suárez Mascareño, A., Figueira, P., et al. 2022, *A&A*, 658, A115  
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, 616, A1  
 Gilbert, E. A., Barclay, T., Schlieder, J. E., et al. 2020, *AJ*, 160, 116  
 Gilbert, E. A., Vanderburg, A., Rodríguez, J. E., et al. 2023, *ApJL*, 944, L35  
 Gillon, M., Triard, A. H. M. J., Demory, B.-O., et al. 2017, *Natur*, 542, 456  
 Goldreich, P., & Tremaine, S. 1980, *ApJ*, 241, 425  
 Hansen, B. M. S., & Murray, N. 2012, *ApJ*, 751, 158  
 Hansen, B. M. S., & Murray, N. 2013, *ApJ*, 775, 53  
 Hansen, B. M. S., & Murray, N. 2015, *MNRAS*, 448, 1044  
 Hayashi, C. 1981, *PThPS*, 70, 35  
 He, M. Y., & Ford, E. B. 2022, *AJ*, 164, 210  
 He, M. Y., & Weiss, L. M. 2023, *AJ*, 166, 36  
 Hellary, P., & Nelson, R. P. 2012, *MNRAS*, 419, 2737  
 Herur, A. N., Tajmohamed, R., & Ponsam, J. G. 2022, in 2022 Int. Conf. on Computer Communication and Informatics (ICCCI)  
 Hsu, D. C., Ford, E. B., Ragozzine, D., & Ashby, K. 2019, *AJ*, 158, 109  
 Ida, S., & Lin, D. N. C. 2010, *ApJ*, 719, 810  
 Izidoro, A., Ogihara, M., Raymond, S. N., et al. 2017, *MNRAS*, 470, 1750  
 Kajtazi, K., Petit, A. C., & Johansen, A. 2023, *A&A*, 669, A44  
 Kirkpatrick, J. D., Henry, T. J., McCarthy, D. W. J., et al. 1991, *ApJS*, 77, 417  
 Kley, W., & Nelson, R. P. 2012, *ARA&A*, 50, 211  
 Kopparapu, R. K. 2013, *ApJL*, 767, L8  
 Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, *ApJ*, 765, 131  
 Li, Y., Stefansson, G., Robertson, P., et al. 2017, *RNAAS*, 1, 49  
 Lillo-Box, J., Figueira, P., Leleu, A., et al. 2020, *A&A*, 642, A121  
 Lin, D. N. C., & Papaloizou, J. 1979, *MNRAS*, 186, 799  
 Lissauer, J. J., Marcy, G. W., Bryson, S. T., et al. 2014, *ApJ*, 784, 44  
 Lykawka, P. S., & Ito, T. 2019, *ApJ*, 883, 130  
 Mah, J., Brasser, R., Woo, J. M. Y., Bouvier, A., & Mojzsis, S. J. 2022, *A&A*, 660, A36  
 Masset, F. S., Morbidelli, A., Crida, A., & Ferreira, J. 2006, *ApJ*, 642, 478  
 Mayor, M., Pepe, F., Queloz, D., et al. 2003, *Msngr*, 114, 20  
 Muirhead, P. S., Dressing, C. D., Mann, A. W., et al. 2018, *AJ*, 155, 180  
 Mulders, G. D., Mordasini, C., Pascucci, I., et al. 2019, *ApJ*, 887, 157  
 Mulders, G. D., Pascucci, I., Apai, D., & Ciesla, F. J. 2018, *AJ*, 156, 24  
 NASA Exoplanet Archive 2019, Kepler Objects of Interest Cumulative Table, 2021-01-21 16:17, IPAC, doi:10.26133/NEA4  
 Nutzman, P., & Charbonneau, D. 2008, *PASP*, 120, 317  
 Ogihara, M., Kokubo, E., Suzuki, T. K., & Morbidelli, A. 2018, *A&A*, 615, A63  
 Ormel, C. W., Liu, B., & Schoonenberg, D. 2017, *A&A*, 604, A1  
 Quintana, E. V., Barclay, T., Raymond, S. N., et al. 2014, *Sci*, 344, 277  
 Raymond, S. N., Barnes, R., & Mandell, A. M. 2008, *MNRAS*, 384, 663  
 Raymond, S. N., & Cossou, C. 2014, *MNRAS*, 440, L11  
 Rodríguez, J. E., Vanderburg, A., Zieba, S., et al. 2020, *AJ*, 160, 117  
 Rowe, J. F., Bryson, S. T., Marcy, G. W., et al. 2014, *ApJ*, 784, 45  
 Schlichting, H. E. 2014, *ApJL*, 795, L15  
 Suárez Mascareño, A., Faria, J. P., Figueira, P., et al. 2020, *A&A*, 639, A77  
 Suárez Mascareño, A., González-Álvarez, E., Zapatero Osorio, M. R., et al. 2023, *A&A*, 670, A5  
 Sun, Z., Ji, J., Wang, S., & Jin, S. 2017, *MNRAS*, 467, 619  
 Tenenbaum, P., Jenkins, J. M., Seader, S., et al. 2013, *ApJS*, 206, 5  
 Terquem, C., & Papaloizou, J. C. B. 2007, *ApJ*, 654, 1110  
 Ward, W. R. 1997, *Icar*, 126, 261  
 Weidenschilling, S. J. 1984, *Icar*, 60, 553  
 Winn, J. N., & Fabrycky, D. C. 2015, *ARA&A*, 53, 409  
 Yu, T. Y. M., Hansen, B., & Hasegawa, Y. 2023, *MNRAS*, 523, 3569  
 Zechmeister, M., Dreizler, S., Ribas, I., et al. 2019, *A&A*, 627, A49  
 Zhu, W., & Dong, S. 2021, *ARA&A*, 59, 291