

Framework for the architecture of exoplanetary systems

I. Four classes of planetary system architecture[★]

Lokesh Mishra^{1,2}, Yann Alibert¹, Stéphane Udry², and Christoph Mordasini¹

¹ Institute of Physics, University of Bern, Gesellschaftsstrasse 6, 3012 Bern, Switzerland
e-mail: exomishra@gmail.com

² Geneva Observatory, University of Geneva, Chemin Pegasi 51b, 1290 Versoix, Switzerland

Received 10 April 2022 / Accepted 22 November 2022

ABSTRACT

We present a novel, model-independent framework for studying the architecture of an exoplanetary system at the system level. This framework allows us to characterise, quantify, and classify the architecture of an individual planetary system. Our aim in this endeavour is to generate a systematic method to study the arrangement and distribution of various planetary quantities within a single planetary system. We propose that the space of planetary system architectures be partitioned into four classes: similar, mixed, anti-ordered, and ordered. We applied our framework to observed and synthetic multi-planetary systems, thereby studying their architectures of mass, radius, density, core mass, and the core water mass fraction. We explored the relationships between a system's (mass) architecture and other properties. Our work suggests that: (a) similar architectures are the most common outcome of planet formation; (b) internal structure and composition of planets shows a strong link with their system architecture; (c) most systems inherit their mass architecture from their core mass architecture; (d) most planets that started inside the ice line and formed in-situ are found in systems with a similar architecture; and (e) most anti-ordered systems are expected to be rich in wet planets, while most observed mass ordered systems are expected to have many dry planets. We find, in good agreement with theory, that observations are generally biased towards the discovery of systems whose density architectures are similar, mixed, or anti-ordered. This study probes novel questions and new parameter spaces for understanding theory and observations. Future studies may utilise our framework to not only constrain the knowledge of individual planets, but also the multi-faceted architecture of an entire planetary system. We also speculate on the role of system architectures in hosting habitable worlds.

Key words. planetary systems – planets and satellites: formation – planets and satellites: dynamical evolution and stability – planets and satellites: fundamental parameters

1. Introduction

Over the last 25 yr, our knowledge of exoplanetary astrophysics has improved dramatically. While the first decade was marked by sensational discoveries of individual exoplanets (e.g. [Vidal-Madjar et al. 2003](#); [Santos et al. 2004](#); [Bouchy et al. 2005](#); [Udry et al. 2007](#); [Kalas et al. 2008](#); [Charbonneau et al. 2009](#); [Snellen et al. 2010](#)), we are now in an age of population-level exoplanetary statistics (for a recent review, see [Zhu & Dong 2021](#)). We now know that (statistically) almost every star hosts a planet and one in two Solar-like stars host a rocky planet in their habitable zone ([Hsu et al. 2019](#); [Bryson et al. 2021](#)). Moreover, many exoplanet-hosting stars have multiple planets orbiting them.

The arrangement of multiple planets and the collective distribution of their physical properties around host star(s) characterises the architecture of a planetary system ([Mishra et al. 2021](#)). Exoplanets in some multi-planetary systems are thought to behave like ‘peas in a pod’ ([Lissauer et al. 2011](#); [Ciardi et al. 2013](#); [Millholland et al. 2017](#); [Weiss et al. 2018](#)). The peas in a pod trend consists of the following correlations: size, whereby adjacent exoplanets are either similar or ordered in size (i.e. the outer planet is larger); mass, whereby adjacent exoplanets are

either similar or ordered in mass; spacing, whereby for a system with three or more planets, the spacing between an adjacent pair of exoplanets is similar to the spacing between the next consecutive pair; packing, whereby smaller planets tend to be packed together closely and larger planets are in wider orbital configurations.

While the statistical method used by [Weiss et al. \(2018\)](#) has been debated ([Zhu 2020](#); [Murchikova & Tremaine 2020](#); [Weiss & Petigura 2020](#)), support for the astrophysical nature of the peas in a pod correlations (as opposed to emerging from detection biases) has emerged from theoretical studies and numerical simulations ([Adams 2019](#); [Adams et al. 2020](#); [He et al. 2019, 2021](#); [Mulders et al. 2020](#)). In particular, [Mishra et al. \(2021\)](#) reproduced the observations from [Weiss et al. \(2018\)](#) using a model of planet formation and evolution (the Bern Model [Emsenhuber et al. 2021a,b](#)) and a model for the detection biases of a Kepler-like transit survey (using KOBE). We showed that when nature's underlying exoplanetary population (consisting of detected and undetected exoplanets) resembles peas in a pod, then a population of transiting exoplanets will have correlations that are consistent with those found by [Weiss et al. \(2018\)](#). In addition, [Mishra et al. \(2021\)](#) suggested that the four trends are not independent of each other. The size correlations seem to emerge from the mass correlations, while the mass and packing trends could combine to give rise to the spacing trend. The peas in a pod trends are amenable to a unification.

[★] Catalogue of observed planetary systems used in this work (full Table 1) is only available online at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/670/A68>

Most of the current studies on this topic utilise statistical correlation coefficients at the population level, that is, the correlation is measured for adjacent planetary pairs from several planetary systems. While useful in terms of testing the existence (or otherwise) of architecture trends, these coefficients may have limited utility for analysing the architecture of a single planetary system. Being statistical in nature, a reliable estimate of these coefficients requires large datasets – which seems difficult for a single system. Although there are some planetary system-level studies (Kipping 2018; Alibert 2019; Mishra et al. 2019; Gilbert & Fabrycky 2020; Bashi & Zucker 2021, discussed in Sect. 3.1), the current literature lacks a prescription for uniformly assessing the multi-faceted architectures of several quantities (e.g. mass architecture, radius architecture, or eccentricity architecture) for a single planetary system.

We seek a framework that allows us to characterise the architecture of an individual planetary system. Our motivations for developing such a framework arise from questions related to: formation, such as the extent to which a system’s architecture is shaped by initial conditions (i.e. the environment in and around the star and protoplanetary disk formation regions; Jin & Li 2014; Safsten et al. 2020); evolution, the role of physical processes such as orbital migration or giant impacts in shaping the final architecture of planetary systems (Mulders et al. 2020); identification, which particular stars host planets that resemble peas in a pod, and, in particular, whether the planets in systems like TOI-178 (Leleu et al. 2021), Trappist-1 (Agol et al. 2021), or 55 Cancri (Bourrier et al. 2018) show mass/size similarities; other architectures, we know that there are many planetary systems that do not follow the peas in a pod architecture (e.g. the Solar System). Overall, it is not obvious how the architecture of any individual planetary system should be uniformly assessed.

In this series of papers, we propose a framework for examining the architecture of planetary systems at the system level. The philosophy behind system level analysis is to consider the entire planetary system as a single unit of a physical system. This framework allows us to not only quantify, compare, and investigate a system’s architecture, but also offers some unexpected benefits. As it turns out, the framework allows for a conceptually intuitive partitioning of the space of possible architectures. We label the four classes of planetary system architectures as: similar, ordered, anti-ordered, and mixed. In this way, our work extends the trends initiated by the notion of peas in a pod architecture. Furthermore, we verify the unification of the peas in a pod correlations proposed in Mishra et al. (2021). We find that, Similar architectures are the most common type of planetary system architectures and their high occurrence explains why the intra-system radius uniformity was already observable from the first four months of *Kepler* data (Lissauer et al. 2011).

Our framework engenders novel questions. For instance, if nature produces distinct classes of architecture in multi-planetary systems, then what is the frequency or occurrence rates of these architecture classes? How does the occurrence of an architecture class depend on stellar and protoplanetary disk environment? How does the architecture of a system evolve over time? What is the role of stellar evolution, protoplanetary disk interactions, and planet formation in shaping the final architecture? How is a planet’s internal composition related to the system’s architecture? Or does the ability of a planet to host life depends on the architecture of the planetary system? In this series of papers, we explore these questions. Although the number of multi-planetary systems is low today, this may change in the next few decades. Thanks to large survey missions such as PLATO (Rauer et al. 2014), *Gaia* (Gaia Collaboration

2016), TESS (Ricker et al. 2015), LIFE (Quanz et al. 2022), and others, the growing number of known multi-planetary systems will allow for a better understanding to emerge. We hope our work encourages observers to dedicate more observation time to detecting planets within a known planetary system, that is, in finding multi-planetary systems.

The architecture classification scheme proposed in this paper is a model-independent framework. To demonstrate our classification framework and explore its consequences, we applied our framework to simulated planetary systems. To illustrate our framework on real systems, we also applied our framework to observed exoplanetary systems. We emphasise that while the results emerging from the application of our framework on these datasets may suffer from some limitations (arising from theoretical modelling or detection biases for observed systems); however, the concept of our architecture classification scheme, being model-independent, does not share these limitations. In this paper, we present the catalogues of planetary systems we apply our framework to in Sect. 2, along with a newly curated catalogue of observed exoplanetary systems and simulated planetary systems, using the Bern Model. We introduce our framework in Sect. 3. In Sect. 4, the characteristics of the architecture classes are discussed. We explore the link between the internal composition of planets and the system architecture class in Sect. 5. Then, in Sect. 6, we speculate on how habitability could depend on the architecture of planetary systems. Our conclusions are given in Sect. 7.

In a companion paper, we investigate the formation pathways, i.e. the role of initial conditions and physical processes in shaping the final architecture (Mishra et al. (2023) referred to as Paper II). Our work demonstrates that the processes of planet formation and evolution are imprinted on the entire system-level architecture. We find that protoplanetary disks with low solid-mass give rise to planetary systems endowed with a mass similarity. On the other hand, massive disks and high metallicity often lead to mass Ordered, Anti-Ordered, or Mixed system architectures. Planet-planet and planet-disk interactions play a decisive role in shaping these three architectures.

2. Catalogues

2.1. Theoretical dataset: Bern Model

In this series of works, we demonstrate our architecture framework by analysing the architecture of synthetic planetary systems. These systems were numerically computed using the Generation III Bern Model of planet formation and evolution (Emsenhuber et al. 2021a,b) that is based on the core-accretion paradigm of planet formation (Pollack et al. 1996; Alibert et al. 2004, 2005). The model follows the growth of protoplanetary embryos embedded in a protoplanetary disk of gas and solids around a solar-type star. A diverse range of physical processes are simultaneously occurring and coherently computed in this 1D star-disk-embryo system. These include: stellar and disk physics (evolution of and interaction between star and viscous disk, condensation of volatile and refractory species, etc.), planetary formation physics (accretion of planetesimals and gases, internal structure calculations, etc.), and additional physics (orbital and tidal migration, planet-planet N -body interactions, planet-disk interactions, atmospheric escape, deuterium fusion, etc.). We describe these physical processes in Appendix A and a descriptive summary of these processes is provided in Mishra et al. (2021, in particular, Fig. 1 and Sect. 2, 3, and Appendix A). More details can also be found in Emsenhuber et al. (2021a,b).

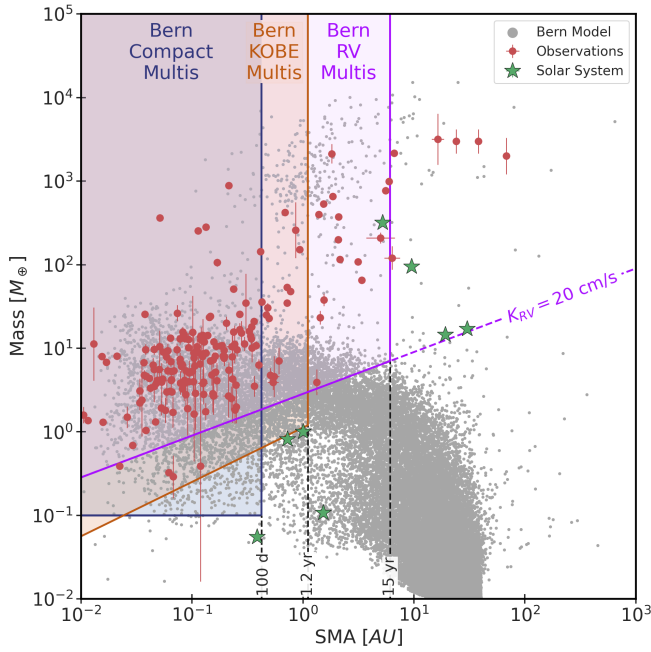


Fig. 1. Mass-distance diagram. This figure shows the masses and the distances of planets in all catalogues used in this study. Shaded regions show the parameter space spanned by synthetic planets observed via radial velocity surveys (Bern RV Multis), transit surveys (Bern KOBE Multis), and ongoing missions (Bern Compact Multis). The parameter space for Bern KOBE Multis has been mapped from its original radius-period plane.

We synthesised 1000 planetary systems, each starting with 100 lunar mass protoplanetary embryos, wherein the following initial conditions were varied: mass of protoplanetary gas disk, photo-evaporation rate, dust-to-gas ratio, disk inner edge, and the starting location of embryos. In Fig. 1, we show all synthetic planets on the mass-distance diagram. For each synthetic planetary system failed embryos, objects with mass less than $0.1 M_{\oplus}$, were removed from further analysis¹.

Three observationally motivated catalogues were prepared from the synthetic dataset. This allowed us to facilitate a comparison of the architecture from observed planetary systems with the synthetic planetary systems and to make predictions. The parameter space spanned by the planets in these catalogues is shown in Fig. 1. These catalogues are as follows:

Bern RV Multis. We assume a radial velocity (RV) survey which can find planets with periods ≤ 15 yr and semi-amplitude $K_{RV} \geq 20 \text{ cm s}^{-1}$. These numbers are motivated by (a) long-running RV surveys such as the HARPS survey (Mayor et al. 2003, 2011) and the California Legacy Survey (Rosenthal et al. 2021; Fulton et al. 2021); (b) current precision achieved by ESPRESSO (Lillo-Box et al. 2021; Netto et al. 2021); and (c) making predictions for future RV surveys. Such RV detectable synthetic planetary systems with four or more planets form the Bern RV Multis catalogue, which includes 3828 planets around 565 stars.

¹ As long as the mass threshold for failed embryos is kept under $0.1 M_{\oplus}$, the results presented in this paper are not sensitive to the threshold limit. We removed these small objects since they (a) failed to grow as massive planets, (b) are insignificant to the dynamical evolution of the system, and (c) are currently unobservable in exoplanetary systems. All results arising from the Bern RV Multis, Bern KOBE Multis, and Bern Compact Multis are insensitive to these failed embryos.

Bern KOBE Multis. We assume a *Kepler*-like transit survey which continuously observes 2×10^5 stars for 3.5 yr (Thompson et al. 2018). A planet which transits three or more times and produces a transit S/N of 7.1 or more is considered detectable. The reliability and completeness of such a survey is replicated and those synthetic planets which would have been vetted as ‘planetary candidates’ by the *Kepler* Robovetter (Thompson et al. 2018), are kept. Such transiting synthetic planetary systems with four or more planets form the Bern KOBE Multis catalogue. KOBE was developed and introduced in Mishra et al. (2021). There are 6 715 planets around 1283 stars in this catalogue.

Bern Compact Multis. Ongoing transit missions such as CHEOPS and TESS have been successful in characterising compact multi-planetary systems, such as TOI-178 (Leleu et al. 2021) and TOI-561 (Lacedelli et al. 2021). Inspired by these discoveries, we investigated the architecture of compact planetary systems simulated by the Bern Model. Our aim is to understand the architecture and make predictions for such systems based on the core-accretion paradigm (Pollack et al. 1996; Alibert et al. 2004, 2005). All planets with periods of ≤ 100 d and masses of $\geq 0.1 M_{\oplus}$ were retained. Synthetic planetary systems, in this parameter space, with four or more planets form the Bern Compact Multis catalogue, with 2412 planets around 400 stars included.

2.2. Observational dataset: A new catalogue

To demonstrate our framework on observed exoplanetary systems, we have curated a new catalogue of known multi-planetary systems². A salient feature of this catalogue (and the philosophy behind this work) is its focus on considering planetary systems as a single unit of a physical system. Unlike focussing on individual exoplanets or a single detection technique, our aim is to study the planetary system as a whole. There are two serious challenges to this endeavour. Firstly, the biases present in detection methods tend to prevent a complete, reliable picture of an exoplanetary system from emerging (either via undetected or mischaracterised planets). Secondly, detecting planets on long orbital periods requires long-term, repeated observations, which is considerably challenging. We hope that upcoming missions and future surveys can mitigate these difficulties.

We included a planetary system in our catalogue if: (a) it has at least four known planets and (b) masses are available for at least four planets. For example, Kepler-33, a five planet system, is included because mass measurements are available for four of its planets³. The criterion of requiring minimum four planets emerges due to (a) the requirement for enough planets for adequately characterising the architecture and (b) because for systems with lower number of planets, it is perhaps difficult to uniformly assess whether the low multiplicity is an outcome of natural processes or detection biases. To keep the comparison between observations and theory uniform, all catalogues in this series of works only consider planetary systems with four or more planets. The architecture framework can, however, handle two- or three-planet systems as well. To make this catalogue useful to the wider community and enable future studies, we gathered several key stellar and exoplanetary properties. For host stars, we report the mass, radius, luminosity, effective temperature, metallicity, age, and distance, along with their

² The catalogue was last updated in April 2021.

³ For this study, the distinction between mass and minimum mass is ignored.

Table 1. Observed multi-planetary systems: There are 41 planetary systems with 194 planets in this catalogue.

Stellar parameters								
Hostname	Multiplicity	$M_\star[M_\odot]$	$R_\star[R_\odot]$	$L_\star[L_\odot]$	$T_{\text{eff}}[K]$	[Fe/H]	Age [Gyr]	Distance [pc]
Sun	8	1	1	1	5,772	0	04.6 ± 0.1	0
Trappist-1	7	0.1 ± 0.002	0.1 ± 0.001	5.53e-04	2,566 ± 026	+0.04 ± 0.08	07.6 ± 2.2	012.0
TOI-178	6	0.7 ± 0.03	0.7 ± 0.01	0.1 ± 01.08	4,316 ± 070	-0.23 ± 0.05	07.1 ± 6.1	062.7
HD 10180	6	1.1 ± 0.05	1.1 ± 0.04	1.5 ± 00.02	5,911 ± 019	+0.08 ± 0.01	04.3 ± 0.5	039.0
HD 219134	6	0.8 ± 0.03	0.8 ± 0.005	0.3 ± 00.01	4,700 ± 020	+0.11 ± 0.04	11.0 ± 2.2	006.5
Planetary parameters								
Hostname	Planet	$M_p[M_\oplus]$	$R_p[R_\oplus]$	$a_p[AU]$	e	i [°]	min. M_p	
Sun	(m,v,e,m, j,s,u,n)	(00.055 ± 00.000,)	(0.383 ± 0.000,)	(0.387 ± -,)	(0.21 ± -,)	(7.00 ± -,)	(F,F,F,F,...)	
		(00.815 ± 00.000,)	(0.949 ± 0.000,)	(0.723 ± -,)	(0.01 ± -,)	(3.39 ± -,)		
			
Trappist-1	(b,c,d,e, f,g,h)	(01.374 ± 00.069,)	(1.116 ± 0.014,)	(0.012 ± 0.000,)	(0.01 ± 0.00,)	(89.73 ± 0.17,)	(F,F,F,F,...)	
		(01.308 ± 00.056,)	(1.097 ± 0.014,)	(0.016 ± 0.000,)	(0.01 ± 0.00,)	(89.78 ± 0.12,)		
			
TOI-178	(b,c,d, e,f,g)	(01.500 ± 00.440,)	(1.152 ± 0.073,)	(0.026 ± 0.001,)	(- ± -,)	(88.80 ± 1.30,)	(F,F,F,F,...)	
		(04.770 ± 00.680,)	(1.669 ± 0.114,)	(0.037 ± 0.001,)	(- ± -,)	(88.40 ± 1.60,)		
			
HD 10180	(c,d,e, f,g,h)	(13.222 ± 00.445,)	(- ± -,)	(0.064 ± 0.001,)	(0.07 ± 0.03,)	(- ± -,)	(T,T,T,T,...)	
		(12.014 ± 00.699,)	(- ± -,)	(0.129 ± 0.002,)	(0.13 ± 0.05,)	(- ± -,)		
			
HD 219134	(b,c,f, d,g,h)	(04.620 ± 00.140,)	(1.544 ± 0.059,)	(0.039 ± -,)	(0.00 ± -,)	(85.19 ± 0.13,)	(F,F,T,T,...)	
		(04.230 ± 00.200,)	(1.458 ± 0.048,)	(0.065 ± -,)	(0.06 ± 0.04,)	(87.38 ± 0.10,)		
			

Notes. Only the first five rows are shown here. The entire table is available at the CDS. CDS version includes additional identification columns: KIC ID, TIC ID, and *Gaia* ID. Missing information is indicated by ‘-’. References for individual systems are given in Appendix B.

identification numbers (when available) in the *Kepler* Input Catalogue (KIC), TESS Input Catalogue (TIC), and *Gaia* ID. For planets, we report mass or minimum mass, radius, semi-major axis, eccentricity, and inclination. In a conservative approach, errors (reported when possible) are the maximum of the upper and lower error bounds available in the literature. When multiple publications reported planetary parameters, a more recent publication was preferred. When a single publication reported parameters for all planets in a system, then such a consistent set of solution was given preference (e.g. GJ 676 A or Kepler-11). For stellar parameters, if a star was included in KIC, then the values from Berger et al. (2020) are reported. Most other stellar parameters come from the TIC (Stassun et al. 2017) or from individual publications.

There are 41 planetary systems that meet our criteria and define our multi-planetary system catalogue (Table 1). With a total of 194 planets in our catalogue, the number of planetary systems with four, five, six, seven, and eight planets is 24, 7, 8, 1, and 1. In this paper, we present the observed planetary systems as they are known today and we do not correct the observations for any detection biases. Instead, to assist in making comparisons with the theory, detection biases will be placed on simulated planetary systems (Sect. 2.1). Figure 1 shows the mass of observed exoplanets as a function of their semi-major axis.

While our observed multi-planetary systems catalogue engenders system-level studies, its current form poses several technical difficulties. Foremost, the number of observations is only forty-one. Secondly, multiple detection methods, such as radial velocity or transits (etc.) were employed to observe these planetary systems. Each observation technique suffers from

certain limitations and detection biases. This implies that the observed systems in our catalogue do not constitute a homogeneous and complete set of observations. These two limitations of the observations catalogue prohibit us from deducing any statistically strong result. Nevertheless, we used the observed systems for (a) exemplifying system-level approach to real planetary systems and (b) using our framework on observations to explore trends in the architecture of observed systems.

Our results from the observed catalogue may be affected by another source of difficulty. There are two systems in our catalogue that host some planets without known mass measurements (Kepler-33 b and Kepler-80 f and g). Since these two systems have at least four planets with known masses, they have been included in our study. However, this does not impact the results of the present study in a drastic way. All three planets in these systems without mass measurements are either the innermost and/or the outermost planets in their respective systems. Therefore, the missing measurements do not have a strong influence on the characterisable mass architecture. The missing measurement may have a strong effect if any planet with unknown mass was in between two planets with known masses.

3. Characterizing architecture: A new framework

3.1. Literature review

We review some approaches from other studies that have tried to capture planetary system-level properties in this section. Kipping (2018) investigated similarity and ordering (of planetary sizes) at

the level of an individual system. Using an entropy based framework on *Kepler* systems, he concludes that initial conditions are inferable from the present-day architecture. As we go on to show in this series, our work not only supports this conclusion, but additionally demonstrates the possible links between initial conditions and final architecture. Although the above-mentioned study considers a similar problem to the one we deal with here, our frameworks differ considerably. Built on step-functions and combinatorics, the aforementioned framework does not take into account the magnitude of variation.

Alibert (2019) proposed a concept of distance between two planetary systems. The Alibert distance captures inter-system differences, whereas our framework quantifies intra-system similarities. The Alibert distance is useful to quantify the similarity (or dissimilarity) between two planetary systems and in unsupervised machine-learning algorithms to find clusters in the space of planetary systems. Bashi & Zucker (2021) recently proposed another concept for distance based on a statistical distance. The ‘weighted’ energy distance is the distance between two planetary systems, with each planet represented on the log-period and log-radius plane, utilising planetary masses (from a mass-radius relationship) as weights. As with the Alibert distance, the Bashi-Zucker distance requires two planetary systems and thus it is not suitable for characterising the global architecture for a single planetary system.

Gilbert & Fabrycky (2020) proposed seven parameters for quantifying the global structure of planetary systems: dynamical mass (ratio of mass in planets to stellar mass), mass partitioning (normalised mass disequilibrium), mass monotonicity (weighted Spearman correlation coefficient), characteristic spacing (average mutual Hill radii), gap complexity, flatness, and multiplicity (n). Of these measurements, mass partitioning and mass monotonicity have close parallels with our framework. The input information required to compute mass partitioning, and monotonicity is exactly the same as the input information for our architecture framework, namely, a set of planetary masses. However, we find that the output displays a curious mix of concepts.

Mass partitioning is zero for a system in which all planets have the same mass. When one planet has some mass and all other planets have negligible mass, the mass partitioning for this system is unity. While this parameter captures the two extreme cases, it is difficult to interpret and employ this measure in cases other than these two extremes. Behaving similarly to a correlation coefficient, mass monotonicity has a range of $[-1,1]$. It is defined as the Spearman correlation coefficient (between mass and distance) multiplied by the mass partitioning (which is weighted by n^{-1}). Although the work of Gilbert & Fabrycky (2020) studies the architecture of planetary systems at the system-level, we seek a framework which can also be used with planetary properties other than mass, such as radius, bulk density, water mass fraction, eccentricities, and so on.

Millholland et al. (2017) and Wang (2017) showed that the peas in a pod pattern reported by Ciardi et al. (2013); Weiss et al. (2018) also extends to planetary masses. Millholland et al. (2017), using planetary masses derived from transit-timing variations, studied the clustering of planets in the mass-radius plane and found that the sum of distances (in the log mass-size space) between adjacent planets of real systems is much smaller than a bootstrapped randomised population. Based on a set of 29 RV observed systems, Wang (2017) infer two types of planetary systems. Planetary systems with masses of $\lesssim 30 M_{\oplus}$ show intra-system mass uniformity, while systems with masses $\geq 100 M_{\oplus}$ do not follow the peas in a pod pattern – indicating that there are

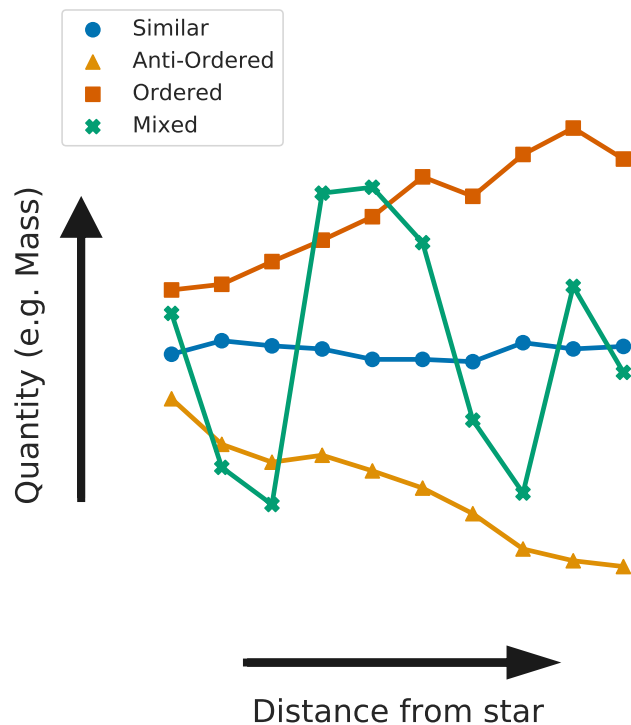


Fig. 2. Classes of architecture. This schematic diagram shows the four architecture classes: similar, anti-ordered, mixed, and ordered. Depending on how a quantity (e.g. mass or size) varies from one planet to another, the architecture of a system can be identified.

only two possibilities for the architecture structure. As we show in this series of works, their hypothesis of only two architecture types is too simple and cannot capture the richness of physics.

3.2. Concept

With our framework, we initially aimed to capture the key aspect about the peas in a pod architecture trends. These trends are correlations between adjacent planets or between consecutive pairs of adjacent planets. We want to capture these ideas at the level of a single planetary system through a unified framework. We do this by studying how a quantity, q_i , (such as mass, size, or period ratio) varies for all planets within a system. Here, i indexes the planets within a system. For all quantities, we adopt an ‘inside-out’ convention, namely, we start with the innermost planet ($q_{i=1}$) and go to the next adjacent planet ($q_{i=2}$), and so on. By comparing how q_i varies for each planet inside-out, we are actually estimating how q_i varies with distance from the host star.

In comparing a quantity, q_i , with distance, four kind of variations emerge. In one scenario, a quantity could show little to no variation. In another case, the value of a quantity may increase with increasing distance or, conversely, the quantity could decrease from one planet to another. Finally, it is also possible for a quantity to not have any clear variations from one planet to another. We identify these four scenarios as the four classes of architectures that can exist at the level of a single planetary system. This idea is depicted in Fig. 2.

Mishra et al. (2021) suggested that the mass correlations could originate from planet-formation physics and the correlations of size and spacing could be derivative. Therefore, we first apply our framework using planetary masses (except in Sects. 5 and 6). As depicted in Fig. 2, when the masses of

all planets within a system are similar to each other, we label the architecture of such systems as ‘similar’. This architecture class corresponds to the peas in a pod architecture reported in observations (Weiss et al. 2018; Millholland et al. 2017). When the masses of planets tend to increase inside-out, the architecture of such systems is labelled ‘ordered’. If the planetary mass tends to decrease from the inner planet to the outer, we label the architecture of these systems as ‘anti-ordered’. Finally, if a large increasing and decreasing variation in the planetary masses is present, we label the architecture of such systems as ‘mixed’. The mixed architecture class is also useful in capturing all other architecture patterns which do not fall under the other three architecture classes. Kipping (2018), for example, has analysed some interesting repeating patterns. By introducing these architecture classes, our framework organises the possibilities for system architecture.

One might wonder, at this point, why introduce such a concept and the ensuing mathematical machinery? While part of this work began as an inspired exploration to categorise our understanding of system architecture, it turns out that there are good physical reasons to pursue this process. As is shown in this and a companion paper, planetary systems that have the same architecture tend to have a host of other properties in common, such as internal structures (core-mass, ice-mass) distributions. Most importantly, systems with a common architecture tend to have same formation pathways, initial conditions, and evolutionary histories. Practically, this means that a quick glance at a system’s architecture may reveal a lot more about its formation scenario.

Our architecture classification framework utilises two quantities – the coefficient of similarity and the coefficient of variation, introduced in Sects. 3.3 and 3.4, respectively. These two coefficients allow us to quantify the conceptual ideas we have presented above. Together, these coefficients define a new space of possibilities for system architectures. In Sect. 3.5, we identify the regions of this architecture space that correspond to the four architecture classes introduced above. As this framework deals with the architecture of multi-planetary system, systems with only one planet are not studied within this framework.

3.3. Coefficient of similarity

The term ‘coefficient of similarity’ is commonly used in the fields studying statistics of ecology and genetics (Gower 1971; Dalirsefat et al. 2009). We borrow the term but develop our own concept and definition. Let q be a planetary quantity such as mass, size, period ratios of adjacent planets, bulk density, eccentricity, and so on⁴. The value of this quantity for the i th planet in a system is denoted by q_i . The coefficient of similarity, C_S , measures how q changes from one planet to another, inside-out. For a system with n planets, it is defined as:

$$C_S(q) = \frac{1}{n-1} \sum_{i=1}^{i=n-1} \left(\log \frac{q_{i+1}}{q_i} \right). \quad (1)$$

There is a clear physical interpretation for $C_S(q)$: the coefficient of similarity measures the average order of magnitude variation in the quantity q from one planet to another. The definition of the coefficient of similarity allows us to map the architecture of

⁴ For quantities which admit zero as a possible value, the coefficient of similarity may become ill-defined. This is a coordinate singularity and can be dealt with an appropriate treatment (see Eq. (4) Sect. 5.4).

a planetary system on a one dimensional axis. When $C_S(q) \approx 0$, then the system’s architecture could imply a similarity in q . When $C_S(q)$ is positive, then planets within a system are ordered in q . Conversely, $C_S(q)$ being negative, implies that the planets are anti-ordered.

We have developed a mathematical formalism to study the sensitivity of the coefficient of similarity. In Appendix C, we derive the limiting values of the coefficient of similarity and present the results here. For example, when the q_i values for all planets in a system are within 10% of each other, then the maximum possible value of $C_S(q)$ is 0.09 (see Eq. (C.10)). For maximum tolerances of 20, 40, 60, and 80%, the maximum possible value of $C_S(q)$ are 0.18, 0.37, 0.60, and 0.95 respectively. In Fig. C.1, we show the dependence of the max $C_S(q)$ on t .

The coefficient of similarity cannot distinguish between two classes of architecture: similar and mixed. Systems which show similarity will have $C_S(q) \approx 0$. However, system with mixed architecture have large increasing and decreasing variations, such that the log of ratios $\frac{q_{i+1}}{q_i}$ cancels itself out. Such systems will also have $C_S(q) \approx 0$. We propose the coefficient of variation to distinguish these two architecture classes. The coefficient of similarity depends on the actual order in which planets exist (inside-out) in a system. As we go on to show, the coefficient of variation does not depend on the ordering of planets in a system.

3.4. Coefficient of variation

The coefficient of variation, C_V , is a standard descriptive statistic used to measure the magnitude of variation in a set of numbers (Katsnelson & Kotz 1957; Sharma et al. 2010; Abdi 2010). It is defined as the ratio of the standard deviation with the mean:

$$C_V(q) = \frac{\sigma(q)}{\bar{q}}. \quad (2)$$

The coefficient of variation is a positive quantity. When all q_i have the same value then $C_V(q) = 0$. Planetary systems consisting of planets that have a small (or large) variability in their q_i values will have a small (or large) value of the coefficient of variation. Now, the distinction between systems showing similarity and mixed architecture is clear. While similar systems will have a low value of the coefficient of variation, mixed systems will have a high value of coefficient of variation.

Since this coefficient is a well known statistical measure, there are some derivations for its limit. A classical result from Katsnelson & Kotz (1957) shows that, for a set of n numbers, the maximum value of the coefficient of variation is $\sqrt{n-1}$. However, this result is only a particular case in our setup. In Appendix C, we develop a mathematical formalism to understand the limits of the coefficient of variation and present the results here. When the q_i values for all planets in a system are within 10, 30, 50, 70, and 90% of each other, the absolute theoretical upper limit of $C_V(q)$ is 0.10, 0.31, 0.58, 0.98, and 2.06 respectively. Figure C.1 shows how this upper limit varies with the maximum tolerance, t , for a system.

3.5. Classifying the architectures of planetary systems

We are interested in obtaining a mapping from the scale-invariant coefficients to an architecture class. In Appendix D, we present some considerations that motivate the selection of boundaries between the four classes. The selected boundaries were additionally tested on thousands of mock planetary systems

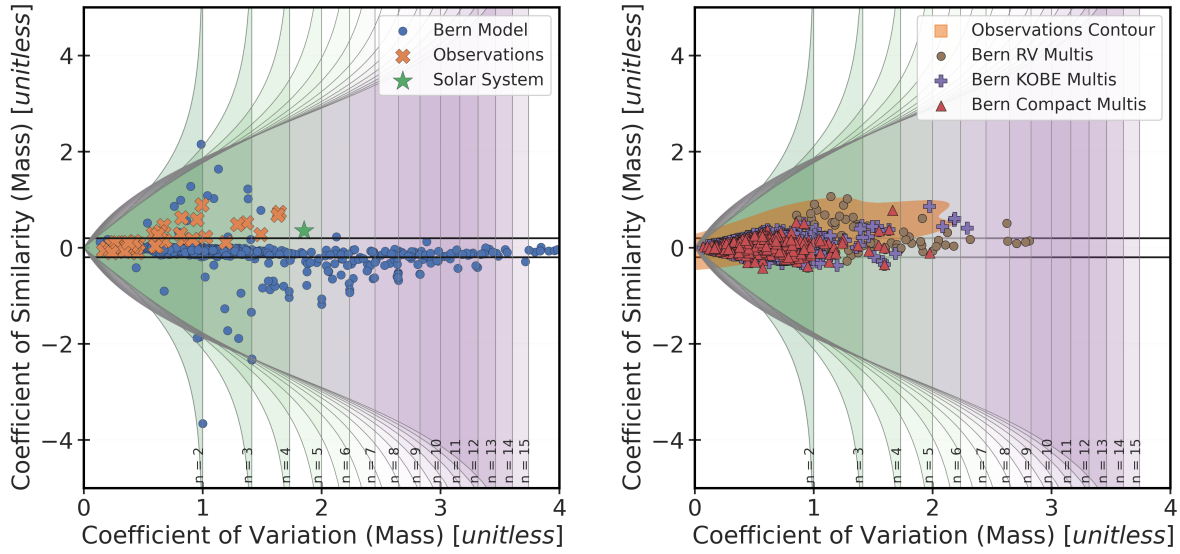


Fig. 3. New parameter space: architectures of planetary systems. Both panels show the coefficient of similarity (mass) as a function of the coefficient of variation (mass). The shaded regions show the allowed parameter space for planetary systems. The white gaps (between two shaded regions) mark the mathematically forbidden regions of this architecture space. Different parts of this parameter space are identified with four architecture classes, in accordance with Eq. (3). Each point corresponds to an individual planetary system. For visual clarity, the shaded and unshaded regions are drawn only for systems hosting up to fifteen planets. Left: planetary systems from the Bern model and observations. Right: synthetically observed systems depicting the detection biases of radial velocity and transit surveys.

to check their ability to correctly classify the four architecture classes. We propose the following boundaries for identifying the architecture class based on planetary masses.

Architecture class	Condition
Anti-ordered	$C_S(M) < -0.2$
Ordered	$C_S(M) > +0.2$
Similar	$ C_S(M) \leq 0.2$ and $C_V(M) \leq \frac{\sqrt{n-1}}{2}$ (3)
Mixed	$ C_S(M) \leq 0.2$ and $C_V(M) > \frac{\sqrt{n-1}}{2}$

A natural (and welcome) outcome of these criteria is that a two-planet system can never have a mixed class architecture. The boundary between similar and mixed class is half the maximum possible value of the coefficient of variation. For the solar system, $C_S(M) = 0.36$ and $C_V(M) = 1.85$. This framework robustly identifies the architecture of the solar system as ordered⁵. This classification is in line with the historic understanding of the solar system architecture: small rocky planets on the inside and giant planets on the outside. If, however, Neptune were replaced with an Earth-like planet, the architecture of the solar system would be classified as mixed. Considering only the inner four planets of the solar system, $C_S(M) = 0.10$ and $C_V(M) = 0.85$, would make the architecture of the inner solar system belong to the similar class. The architecture of the outer four giants in the solar system is anti-ordered and we have $C_S(M) = -0.42$ and $C_V(M) = 1.11$.

Figure 3 shows the $C_S(M)$ versus $C_V(M)$ space for planetary systems from several catalogues. The Bern model planetary

⁵ Even if the masses of each solar system planet were randomly varied within 85% of their original values, the emerging architecture is still ordered. With 1M trials, varying the masses randomly within 90% of their original values lead to ordered (for $\approx 99.45\%$ trials), mixed (for $\approx 0.55\%$ trials), and similar (for $\approx 0.001\%$ trials) architectures.

systems occupy all four regions of this architecture space. Observed planetary systems, however, span only a limited region of this parameter space, given the low multiplicity of observed planetary systems. The architecture space spanned by the observed planetary systems (shaded contour) is in agreement with the synthetically observed planetary systems from Bern Compact Multis, Bern KOBE Multis, and Bern RV Multis.

The architecture for the systems in the synthetically observed catalogue was calculated based only on the planets that were detected (for RV/KOBE) or included (for Bern Compact Multis) in the above-mentioned catalogue. It is theoretically possible for a single Bern model system to exhibit different architectures depending on the planets which are detected or included. The reverse is also true – the architecture of an observed planetary system may change if new planets are discovered or old controversial candidates are rejected. While the ground truth architecture for observations seems elusive, a comparison with synthetic observations can bring forth patterns which are unexpected. With this in mind, we consider the following example.

Detection biases, in both radial velocities and transits, generally disfavour the discovery of less-massive and small planets at larger distances. This implies that anti-ordered architectures are difficult to detect. In fact, we have no known example of a planetary system showing anti-ordered architecture in our observations catalogue. This is surprising for two major reasons: (a) theory suggests their existence: there are several synthetic planetary systems from the Bern Model whose architecture is anti-ordered; (b) theory suggests their discovery: all three synthetically observed catalogues contain some (albeit few) anti-ordered planetary systems. Since the number of systems in our catalogue is too low, we refrain from making any conclusions and, instead, we await the discovery of anti-ordered architectures in the future. However, if such architectures are not found despite considerable efforts, this result will become a strong indicator for shaping our understanding of planet formation.

Another aspect of this new architecture space is the underlying mathematical structure⁶. In Fig. 3, the shaded areas shown regions where a planetary system, with $n \in [2, 15]$ planets, is allowed. A system with two planets, for example, can only occupy the shaded region labelled ‘ $n = 2$ ’. All non-shaded regions (in white – except the shaded regions for 16 or more planets which is not drawn), on this architecture space, is mathematically forbidden. These are parts of the architecture parameter space that no planetary system, irrespective of its configuration, can occupy. This strong result stems from the mathematical limits that were derived for this work (see Sects. 3.3, 3.4, and Appendix C).

For clarity and future convenience, we introduced some terminology to the method. When the architecture framework (i.e. C_S and C_V) is applied on planetary bulk masses, the resulting information tells us the mass architecture of a system, namely, the arrangement and distribution of masses in said system. Similarly, when this framework is applied on radii, it gives us the radius architecture (arrangement and distribution of radii) for the system (Sect. 5.1). Similarly, we can obtain the bulk-density architecture (Sect. 5.2), core-mass architecture (Sect. 5.3), water mass fraction architecture (Sect. 5.4), period-ratio or spacing architecture, eccentricity-architecture, and so on. In this series of papers, we identify a system’s architecture based on its bulk mass architecture. Thus, when a system is said to be similar, we are referring to the similarity in terms of the mass architecture.

4. Characteristics of architecture classes

4.1. General comments

In earlier studies on the peas in a pod architecture, the strength of population-level (i.e. across many planetary systems) trends was quantified using Pearson correlations coefficient (Weiss et al. 2018; Zhu 2020; Chevance et al. 2021; Millholland & Winn 2021; Mishra et al. 2021). The correlation coefficients were calculated using planetary quantities in the log space (i.e. by first taking the \log_{10} of all quantities). This resulted in higher values of the correlation coefficient since quantities have limited range to perambulate in the log space. Consider planetary masses. We calculated the correlation coefficient between the mass of adjacent inner and outer planets in the Bern model population (see Fig. 7 in Mishra et al. 2021). The value of the coefficient is 0.66 in the log space and 0.16 in the linear space. This highlights that the planetary masses are more closely clustered in log than in linear space.

We tested the same correlation for all systems in each architecture class. We expect planetary masses in mixed, ordered, and anti-ordered systems should (by definition) have low correlations. On the other hand, similar class architecture should exhibit a strong correlation. Surprisingly, in log space all architecture classes show strong correlations. The coefficient value is 0.67 for similar class, 0.69 for mixed class, 0.50 for ordered class, and 0.58 for anti-ordered class architectures. However, in the linear space the coefficient values reflects our expectation: 0.61 for the similar class, 0.20 for the mixed class, 0.16 for ordered class,

Table 2. Architecture type of known multi-planetary systems (see Table 1 for catalogue and Fig. 6 for architecture plot).

Hostname	Multiplicity	$C_S(M)$	$C_V(M)$	Architecture class
Solar System	8	+0.36	1.85	Ordered
Trappist-1	7	−0.10	0.45	Similar
TOI-178	6	+0.08	0.46	Similar
HD 10180	6	+0.14	0.66	Similar
HD 219134	6	+0.27	1.49	Ordered
HD 34445	6	+0.17	0.84	Similar
Kepler-11	6	+0.22	1.03	Ordered
Kepler-20	6	+0.00	0.44	Similar
Kepler-80	6	−0.00	0.19	Similar
K2-138	6	+0.03	0.61	Similar
55 Cnc	5	+0.52	1.37	Ordered
GJ 667 C	5	−0.02	0.29	Similar
HD 158259	5	+0.11	0.29	Similar
HD 40307	5	+0.07	0.33	Similar
Kepler-102	5	+0.02	0.41	Similar
Kepler-33	5	+0.46	0.67	Ordered
Kepler-62	5	+0.15	0.68	Similar
HD 20781	4	+0.29	0.59	Ordered
TOI-561	4	+0.33	0.64	Ordered
DMPP-1	4	+0.29	0.81	Ordered
GJ 3293	4	+0.27	0.62	Ordered
GJ 676 A	4	+0.90	0.99	Ordered
GJ 876	4	+0.12	1.20	Mixed
HD 141399	4	+0.06	0.40	Similar
HD 160691	4	+0.63	0.82	Ordered
HD 20794	4	+0.08	0.25	Similar
HD 215152	4	+0.07	0.23	Similar
HR 8799	4	−0.07	0.17	Similar
K2-266	4	+0.03	0.60	Similar
K2-285	4	+0.01	0.31	Similar
Kepler-89	4	+0.17	0.91	Mixed
Kepler-106	4	+0.11	0.26	Similar
Kepler-107	4	+0.13	0.42	Similar
Kepler-223	4	−0.06	0.22	Similar
Kepler-411	4	−0.08	0.34	Similar
Kepler-48	4	+0.74	1.64	Ordered
Kepler-65	4	+0.68	1.63	Ordered
Kepler-79	4	−0.10	0.24	Similar
WASP-47	4	+0.59	0.95	Ordered
tau Cet	4	+0.12	0.37	Similar
HD 164922	4	+0.49	1.29	Ordered

and 0.05 for anti-ordered class. This underscores that strong correlations in the log space may not be indicative of substantive architecture trends. It also shows that our framework is capable of identifying systems in which the ‘peas in a pod’ architecture is discernible even in the linear space.

Figure 4 shows the coefficient of similarity of masses as a function of the total planetary mass in a system for all synthetic planetary systems from the Bern model. This figure shows several key aspects. Firstly, it illustrates the four architecture classes as separate clouds of scattered points strengthening the proposed four classes of planetary system architecture. Secondly, it shows that the architecture framework is scale-invariant, that is, the system architecture is sensitive only to the relative distribution of a quantity – and not its absolute value. For example, while most similar system have $\lesssim 100 M_{\oplus}$ mass in their planets

⁶ Visualizing this structure is easy (not shown). (a) Construct mock planetary systems with masses, for each mock planet, randomly drawn from a uniform distribution with suitable limits. (b) It is suggested to vary the number of planets in these mock systems randomly. (c) Calculate the $C_S(M)$ and the $C_V(M)$ using Eqs. (1) and (2). (d) Plot $C_S(M)$ versus $C_V(M)$ for this mock population. For large number of systems the plot should be symmetric about $C_S(M) = 0$.

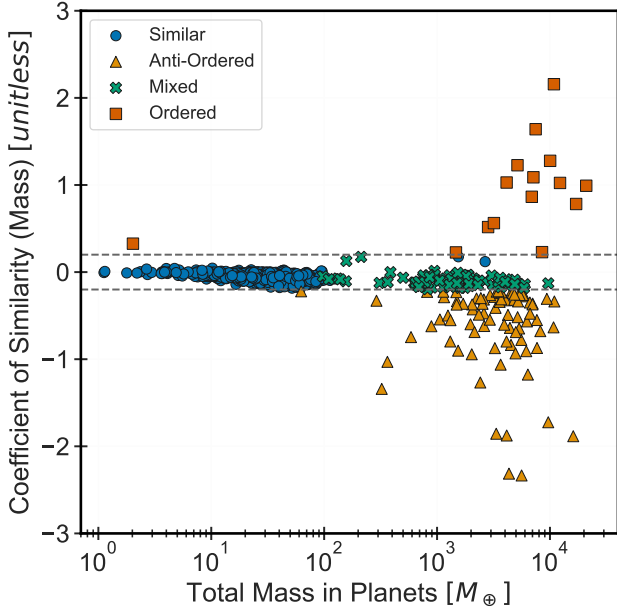


Fig. 4. Four classes of system architecture. The diagram shows the coefficient of similarity for a system as a function of the sum of mass of each planet in a system. Dashed horizontal lines correspond to $C_S = \pm 0.2$. This diagram emphasises the four classes of planetary system architecture, namely: anti-ordered, similar, mixed, and ordered. It also shows that the coefficient of similarity can not distinguish between similar and mixed architectures.

(suggesting a lack of giant planets), there are some similar systems with mass values of $\approx 2000 M_\oplus$ for their planets and host giant planets. Likewise, most ordered systems host giant planets and have $\gtrsim 2000 M_\oplus$ mass in their planets, there is an ordered systems without any giant planets. Also, it illustrates that the coefficient of similarity partitions planetary systems into three groups: anti-ordered, similar and mixed in one group, and ordered. This demonstrates that the coefficient of variation is necessary to distinguish between the similar and mixed systems. Finally, the diagram shows that the architecture class of a system has strong links with the total mass of planets in the system. This hints that there must be general patterns in the formation pathways of systems of the same architecture. This topic is discussed in Paper II, from this series.

For all 41 observed planetary systems in our catalogue, we report their architecture classes in Table 2. The frequency of each architecture class across all catalogues is shown in Fig. 5. Figure 6 shows the architecture of all observed multi-planetary systems in our catalogue. The systems are sorted by their coefficient of similarity values. The figure also shows the four classes of architecture for a few randomly selected synthetic planetary systems. To understand the characteristics of the different architectures, we study the distribution of planetary masses, radii, and semi-major axes as well as the multiplicity distributions. For planetary systems across all catalogues, this is shown in Fig. 7. We describe the characteristics of different architectures in the following subsections. The discussion in the next subsection involves results derived from both observed and synthetic planetary systems. In addition, we present a gallery of mass-distance diagrams showing the four architecture classes in Appendix E.

4.2. Frequency of architecture

Similar systems are the most common architecture classes emerging from simulations, with a frequency of $\approx 80.2\%$. About

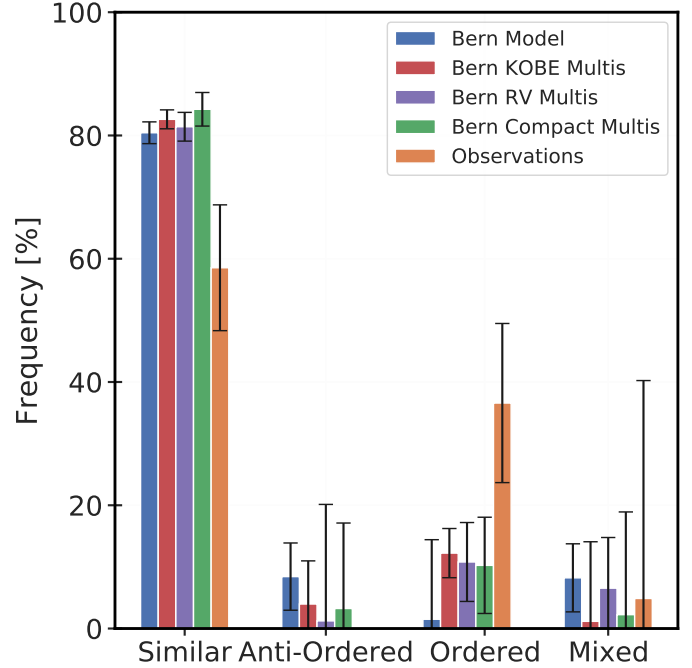


Fig. 5. Frequency diagram for the architecture classes. Currently, there are no known examples of observed planetary systems with anti-ordered architecture. The length of error bars visualises the total number of systems in each bin as: $100/\sqrt{\text{bin counts}}$.

$\approx 8\%$ of synthetic systems show mixed and anti-ordered architectures. Ordered architecture is a rare outcome in simulations ($\approx 1.5\%$). In observations, similar class is the most common architecture ($\approx 59\%$). Fifteen observed exoplanetary systems (out of forty-one) are part of the ordered architecture class ($\approx 37\%$). About $\approx 5\%$ of observed planetary systems show mixed architecture. There are no known examples of observed system with anti-ordered architecture.

Comparing the frequency of architecture classes for observed systems with synthetically observed systems brings out some peculiar features. Firstly, theoretical catalogues seem to suggest that observations should find more similar systems and fewer ordered systems. The frequency of similar (ordered) systems in our observed catalogue is significantly lower (higher). Secondly, while the frequency of mixed systems seems to be in agreement with synthetic observations, this agreement is not statistically significant.

These discrepancies probably arise from the incompleteness prevalent in our observations catalogue. Transit surveys are conducted in a manner which allows the completeness and reliability of these survey to be estimated. The completeness of RV surveys, on the other hand, is very difficult to estimate. Further, the observation techniques used to find the exoplanets in our observations catalogue are heterogeneous, consisting of RV, transits, transit-timing variations, and direct imaging; this complicates the estimation of completeness. The PLATO mission is an upcoming space mission that is equipped to allow for statistical estimates of cosmic occurrence rates of planetary system architecture in our galaxy (Rauer et al. 2014). If more exoplanetary systems are uniformly detected and characterised, then it would be possible to estimate the occurrence rate of the different classes of system architecture. While such a result would constitute an important knowledge about our Universe, it could also become an excellent way of constraining our knowledge of

initial conditions for planetary formation and the physical processes which shape the system architecture. The frequency of architecture class in simulations is a direct consequence of the initial conditions and the physical processes modelled in the Bern model.

4.3. Architecture class: similar

Planetary systems have a similar architecture when all planets in the system have masses that are approximately similar to each other. These planetary systems are the archetypical examples of the peas in a pod trend. There are several well-known planetary systems exhibiting similar architecture, such as Trappist-1 (Agol et al. 2021), TOI-178 (Leleu et al. 2021), Kepler-20 (Buchhave et al. 2016), and so on. This architecture is the most common outcome of planetary formation and is also the most frequent architecture class in our observed catalogue.

Similar systems in the Bern model are composed of several low-mass planets. They tend to have limited diversity in planetary masses when compared with the observed systems. The mass distribution, for similar systems in the Bern model, shows that there are many low-mass ($<1 M_{\oplus}$) planets in these systems. This peak is missing in observations as well as synthetic observations as low mass exoplanets are difficult to observe. This could, however, be remedied in future as current radial velocity spectrographs reach the $\approx 20 \text{ cm s}^{-1}$ precision necessary for discovering exoplanets in the super-Earths and Earths mass range (Lillo-Box et al. 2021; Netto et al. 2021). The radius distribution of similar systems implies that these systems are prominently composed of rocky planets, super-Earths and sub-Neptunes⁷.

The Bern RV Multis show a bimodal planetary distance distribution for similar systems (as well as for mixed and ordered). The approximate location of the gap is 0.28 au or 55 d (for a solar mass star). This bi-modality is not visible in our observed catalogue. Planets in similar and mixed systems in the Bern Model also show a dip around this location. In the Bern Model, inwardly migrating giant planets ($\geq 100 M_{\oplus}$) tend to stop around 0.4 au or 100 d. Inside this region, low-mass planets are populous. We attribute this bi-modality to these two populations of planets. This bi-modality probably arises because planets switch their orbital migration from type I to type II depending on their masses (Emsenhuber et al. 2021a). This bi-modality cannot be seen in Bern Compact Multis because we only include planets with periods less than 100d. For Bern KOBE Multis, the completeness of the *Kepler* mission for large distant planets is poor (see Fig. C.2 in Mishra et al. 2021). However, a dip at this location in Bern KOBE Multis is visible. It would be interesting to see if such a bi-modality is also present in the *Kepler* catalogue. We tested the significance of this bi-modality with Hartigan's dip test (Hartigan & Hartigan 1985). The dip test is suggestive of the bi-modality for the Bern RV Multis and Bern KOBE Multis (p -value < 0.05) and insignificant for the other catalogues.

A system's architecture is sensitive only to the relative distribution of a quantity (such as mass) amongst its planets and not the absolute distribution. HR 8799 offers an example (Morois et al. 2008) as a relatively young system with four directly imaged giant planets. Our framework identifies the architecture of this systems as similar. Most observed similar systems are composed of low-mass planets ($\leq 100 M_{\oplus}$), making HR 8799 a

unique exception. This shows that the architecture framework is sensitive only to the relative variations in the mass. Additionally, there are only two systems (out of 1000) in our simulated catalogue where a similar architecture arises from only giant planets. Even then, these two synthetic systems have only two giant planets much closer to the star than the HR 8799 planets. The Bern Model does not produce many HR 8799-like systems. This suggests that a system with similar architecture made up of only giant planets is probably rare. One possibility could be that systems (e.g. HR 8799) with such architecture are probably difficult to form via core accretion pathway (Konopacky & Barman 2018). Such systems may require additional formation mechanisms such as protoplanetary disk instabilities (Schib et al. 2021; Boley et al. 2010; Kratter et al. 2010).

4.4. Architecture class: mixed

Planetary systems where the planetary masses (inside-out) show broad increasing and decreasing variations have mixed architecture. GJ 876 and Kepler-89 host planetary systems with a mixed class architecture. GJ 876 is an M dwarf low luminous ($\approx 0.01 L_{\odot}$) star hosting four planets with masses between 8–888 M_{\oplus} . The outer three planets are in a Laplace mean-motion resonance (Millholland et al. 2018). Kepler-89, on the other hand, is an early F, highly luminous ($\approx 3.5 L_{\odot}$) star. It hosts a compact four planet system with masses between 10–100 M_{\oplus} . Despite the starkly different stellar properties, the architecture of these two systems is analogous: $C_S(M) = 0.12$ and 0.17 , $C_V(M) = 1.2$ and 0.9 , respectively. While the coefficient of similarity is low for both systems, the coefficient of variation is larger than $\sqrt{3}/2$, which helps us identify the architecture of these systems as mixed class. Indeed, Fig. 6 indicates that this identification is correct.

The frequency of this architecture class in the Bern model is $\approx 8.2\%$. The Bern model's synthetic mixed architecture planetary systems (Fig. 6 right) tend to have numerous Earth-mass planets outside 10 au. This parameter space (mass-distance plane, Fig. 1), however, remains inaccessible to most exoplanet detection techniques. These systems are also composed of super-Earths, sub-Neptunes, Neptunes, and Jovian planets. The bi-modality in distance distribution (discussed before) is prominent for these architectures in Bern RV Multis. We found a Hartigan's dip statistic of 0.03 and p -value of ~ 0.2 (Hartigan & Hartigan 1985).

4.5. Architecture class: anti-ordered

Planetary systems where the planetary mass shows an overall decrease with distance have an anti-ordered architecture. There are no observed examples of this architecture class in our catalogue. The frequency of this architecture class in the Bern model is $\approx 8.4\%$. About $\approx 4\%$ of systems in Bern KOBE Multis, $\approx 3.2\%$ of systems in Bern Compact Multis, and $\approx 1.2\%$ of systems in Bern RV Multis have this architecture. This shows that it is an observationally challenging system architecture to detect. However, even if 1% of observed exoplanetary systems are Anti-Ordered we should already have found about 30–40 such systems. More work is necessary to identify the handful of these systems from the already observed systems. Many currently known single hot Jupiter systems may host additional small, distant, and as yet undetected planets – revealing these potentially anti-ordered systems.

⁷ Throughout this paper, we use planetary classes (e.g. rocky, super-Earths, etc.) from the radius based classification of Kopparapu et al. (2018).

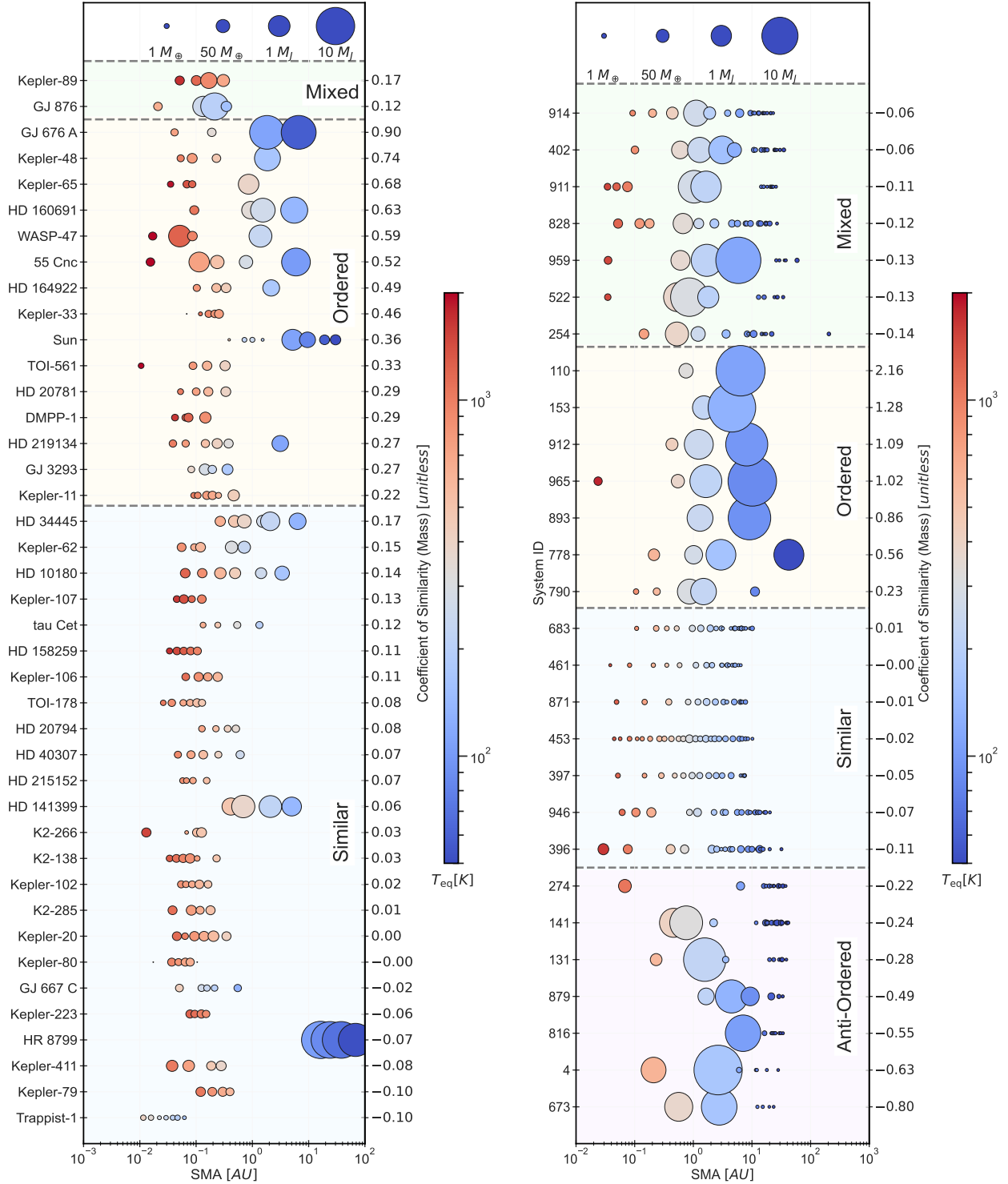


Fig. 6. Architecture plot showing the architecture of observed (left) and randomly selected synthetic planetary systems (right). Each row is for one planetary system and the circles in that row represent planets. The area of the circle encodes planetary mass, and the colour shows the equilibrium temperature. The coefficient of similarity for each system is shown on the right y-axis. The x-axis shows the semi-major axis, which is different for the two panels.

Anti-ordered systems in the Bern Model are mostly composed of low mass planets $\lesssim 5 M_{\oplus}$ and giants $\gtrsim 100 M_{\oplus}$. In the Bern Model, the radius distribution of this architecture class peaks for Rocky and Super-Earths planets. It decreases for sub-Neptunes and Neptunes and then increases again for Jovian planets. Many of the low-mass planets that make up this architecture class are outside 10 au, making their detection very challenging.

The multiplicity distribution shows that these systems tend to have fewer planets than similar or mixed architecture. This is an indication that the formation pathway of these architectures differs considerably from the other two types of architecture. Planets from anti-ordered architectures show a weak distance bi-modality feature (discussed earlier in this work). This is understandable since these architectures consist of massive

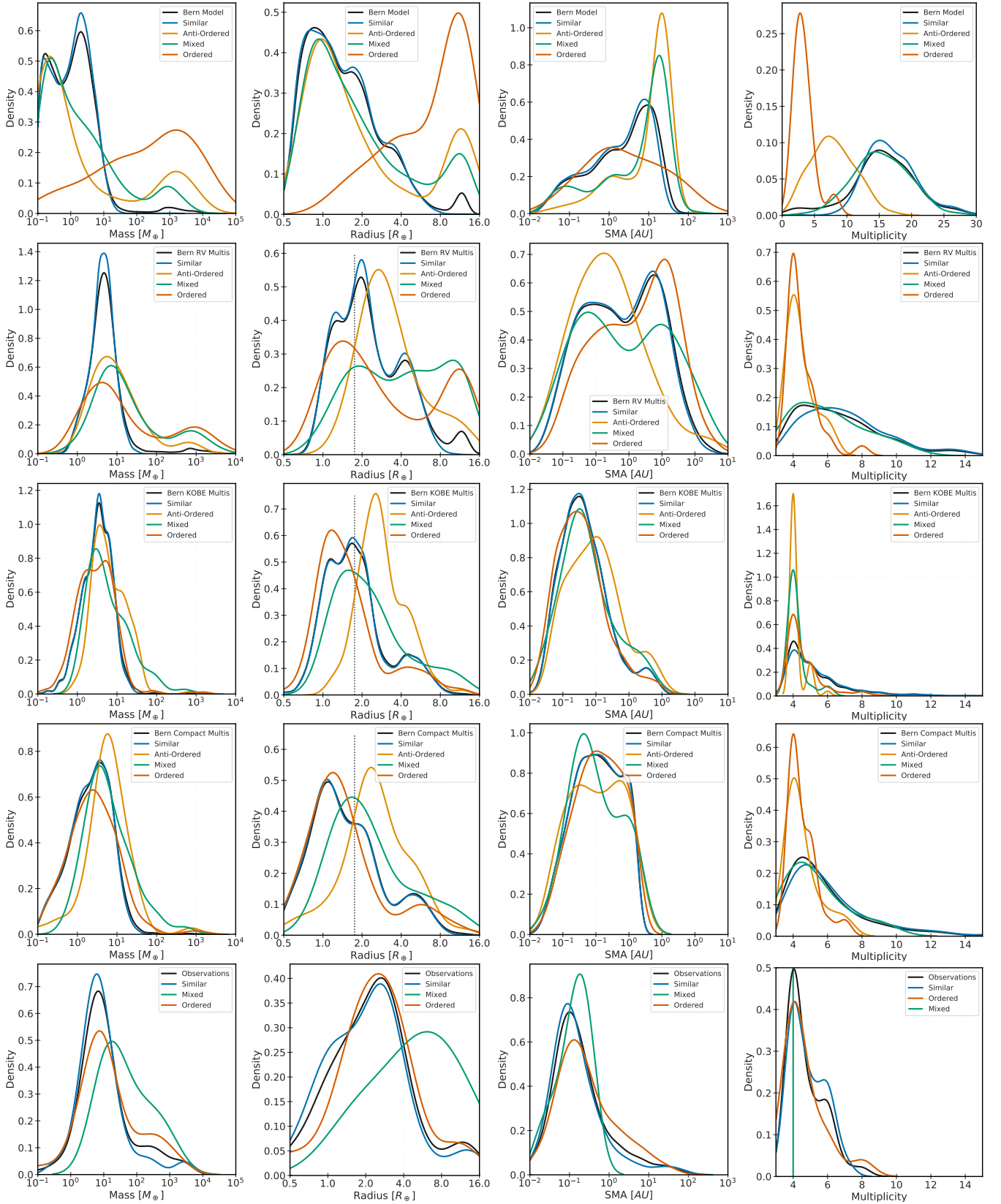


Fig. 7. Characteristics of the architecture classes. These plots show the distribution of various quantities (columns) as function of different catalogues (rows). Left to right: distributions of mass, radius, distance, and multiplicity in the following catalogues (top to bottom): Bern model, Bern RV Multis, Bern KOBE Multis, Bern Compact Multis, and observations. All catalogues are described in Sect. 2. Some notable features from these plots are discussed in Sect. 4. All individual distributions are normalised such that the area under each curve sums to unity. The dotted vertical line in the radius distributions marks $1.75 R_{\oplus}$ – approximately, the location of the well-known gap in the radius distribution (Fulton et al. 2017). Since there are only two mixed systems with the same multiplicity ($n = 4$) in our observations catalogue, a vertical line replaces the density kernel. The Gaussian density kernels in all other cases were estimated using Scott’s rule (Scott 2015).

planets in the inner parts and less massive planets in the outer parts of the system. The distance bi-modality seems to arise from low mass planets (migrating via type I) inside 0.28 au or 55 days and giant planets (migrating via type II) outside 0.28 au or 55 days. This adds further strength in attributing the distance bi-modality to planetary migration.

4.6. Architecture class: ordered

Planetary systems where the planetary masses shows an overall increase with distance have an ordered architecture. The increasing mass may be monotonic (e.g. TOI-561, HD 20781, DMPP-1, HD 160691, HD 164922) or non-monotonic (e.g. the Solar System, Kepler-11, 55 Cnc, Kepler-48, Kepler-65). Ordered architecture is a rare outcome for the Bern model. Observations are generally biased against discovering small and less massive planets which are farther away from their host star. Such biases, however, make ordered systems the second most common architecture class. Fifteen systems in our catalogue exhibit this architecture. Unsurprisingly, the most notable known example of this architecture class is the Solar System.

The mass and radius distributions of ordered architecture in the Bern Model shows considerable difference from other architecture. The mass distribution peaks around $1000 M_{\oplus}$. Most of the Bern model's ordered systems tend to have at least one giant planet. These systems are also composed of sub-Neptunes, Neptunes, and Jovian planets.

5. Internal composition across architecture classes

So far we have seen the new architecture framework (Sect. 3) and some characteristics of the four classes of architecture (Sect. 4). In this section, we study the connection between the bulk mass architecture classes and the internal composition of the planets. This section demonstrates that the same architecture framework can be used to study the multi-faceted nature of planetary system architecture – from bulk mass architecture to density architecture. We study several different aspects of the planetary internal composition: (a) radius architecture (Sect. 5.1); (b) bulk density architecture (Sect. 5.2); (c) Core/Envelope mass architecture (Sect. 5.3); and (d) fraction of volatiles and water ice in core architecture (Sect. 5.4). We explore these connections for planetary systems in the simulated (Bern model) and synthetically observed catalogues (Bern RV Multis, Bern KOBE Multis, Bern Compact Multis). All results in this section are derived from synthetic planetary systems only.

5.1. Radius architecture

Weiss et al. (2018) showed that the size of adjacent exoplanets were similar – coining the phrase ‘peas in a pod’ to describe this architecture. Millholland et al. (2017); Wang (2017) extended these ideas to planetary masses, showing that the masses of adjacent planets are also correlated. In Mishra et al. (2021), we suggested that the peas in a pod trends in terms of size effectively emerge from the mass trends. Here, we attempt to set our assumption on firmer ground.

Figure 8 (top) shows the coefficient of similarity for radii as a function of the coefficient of similarity of masses, for systems with two or more planets. This allows us to compare the system-level radius architecture with the system-level mass architecture. We easily see that most systems seem to follow a linear relationship. The Pearson correlation coefficient is 0.89, indicating a

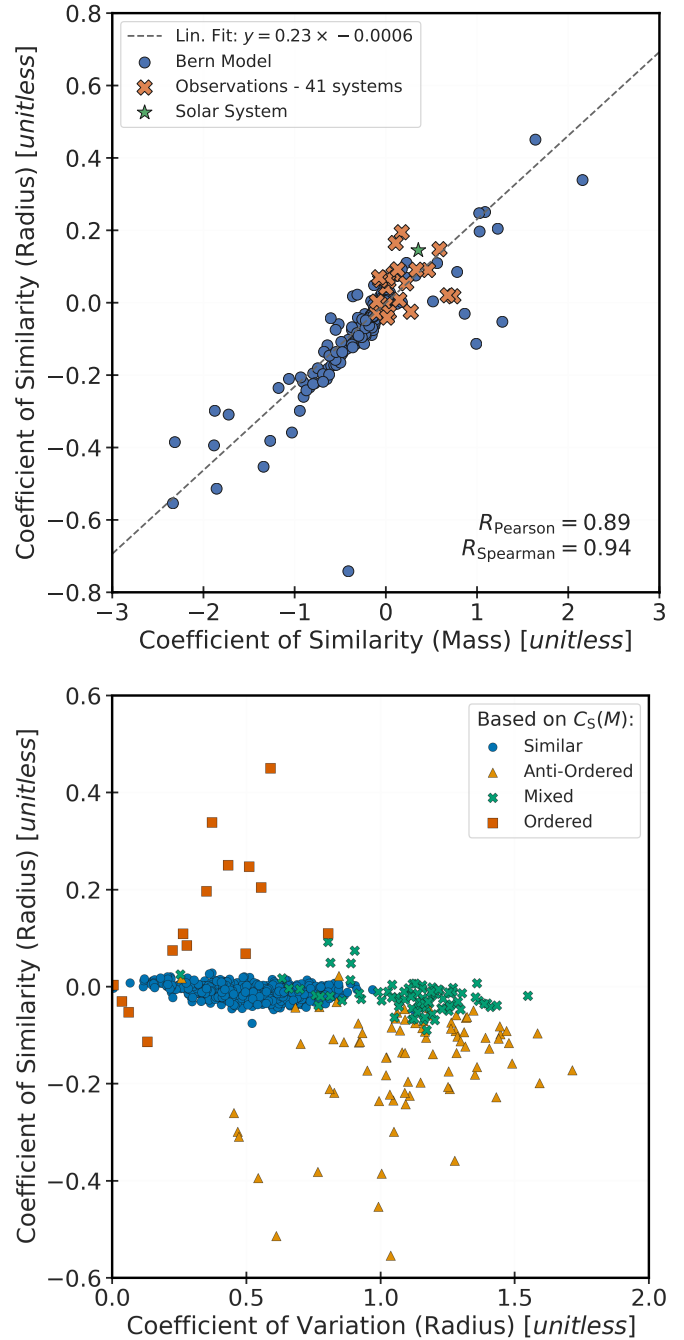


Fig. 8. Radii architecture. Top: the diagram shows the coefficient of similarity of radii as a function of the coefficient of similarity of masses, for synthetic and observed planetary systems. The dashed line shows the corresponding linear fit. Bottom: radius architecture of synthetic planetary systems contrasted with the mass architecture. In the bottom panel, the marker colour and shape indicates the bulk mass architecture of a system and its position on the diagram suggests its radii architecture.

strong positive correlation between the mass and radius architecture. The coefficient value increases to 0.96, when systems with only three or more planets are considered. Since the mass-radius relation is not a bijective function (i.e. one-to-one correspondence), there are some systems that show a strong deviation from the linear relation.

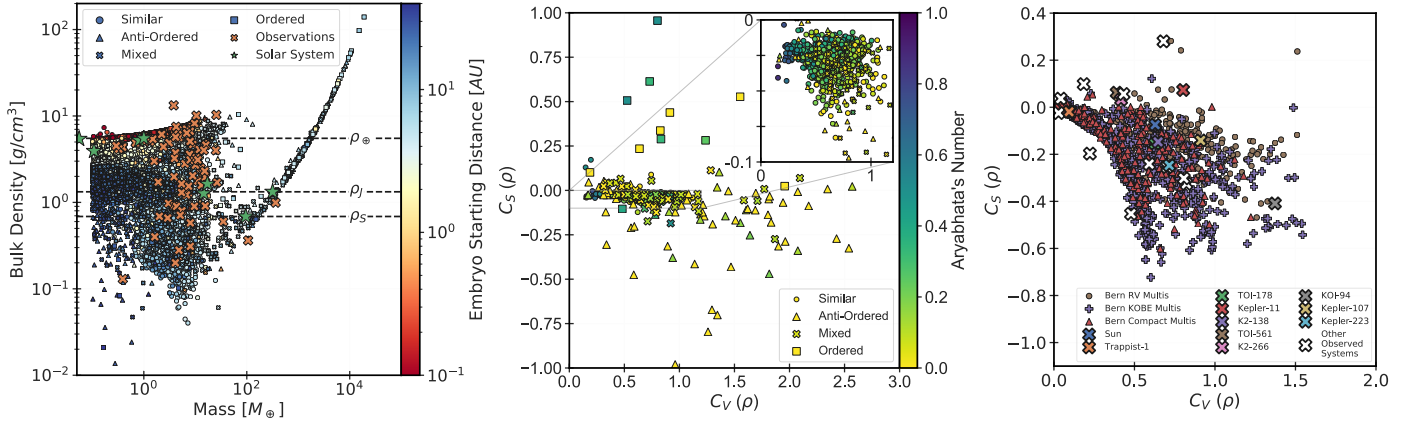


Fig. 9. Density architecture. Left: bulk density of simulated and few observed planets as a function of their mass and starting locations (for synthetic planets). The marker indicates the mass architecture of the system to which a synthetic planet belongs to. Middle: density architecture, of synthetic planetary systems, as seen through the coefficient of similarity versus the coefficient of variation plot. The marker shape and colour indicates their host system mass architecture and the system’s Aryabhata’s number (see Paper II), respectively. Right: density architecture of planetary systems from the simulated observed catalogue and few observed planetary systems.

Figure 8 (bottom) shows the radii architecture for the synthetic planetary systems⁸. This shows that most systems that are ordered (or anti-ordered) in mass are also ordered (or anti-ordered) in terms of radius. The figure also shows that systems which are similar or mixed in mass architecture have $C_S(R) \approx 0$. Systems with mass similarity have lower $C_V(R)$ compared to systems with mass mixture, suggesting that for most systems, the radius architecture closely follows the mass architecture. At the planetary level the radius of a planet is correlated with its mass via the planet’s chemical composition (Lopez & Fortney 2014). Our architecture framework shows that such relationships also exist at the system level. A few mass-ordered systems show similarities in radius. These few systems have the following common features: two mass-ordered giant planets with similar sizes (masses \sim several M_J ’s, and radius $\approx 1 R_J$). This illustrates that while mass architecture and radius architecture are related, they are not always identical.

We conclude that the peas in a pod radius correlations generally arise from the underlying mass architecture. We consider the mass architecture primal because planets, foremost, accrete mass from the protoplanetary disk and, consequently, are characterised by a size that is in accordance with their internal structure.

5.2. Density architecture

Bulk density (or simply density) is a directly measurable quantity which is sensitive to the internal structure of a planet. This makes density an important characteristic for understanding planetary structure. The density of a planet depends on many parameters and many physical processes. For example, a planet’s mass may depend on its accretion history, starting location, amount of material in disk, competition with other planets, and so on. Giant impacts may also affect a planet’s density, as explained in Bonomo et al. (2019). In this section, we study the arrange-

ment and distribution of planetary density around their host star, namely, the density architecture of a system.

Figure 9 (left) shows the density of a planet, simulated via the Bern model, as a function of its mass and starting location. The figure also shows the density of solar system planets and few observed exoplanets (from our catalogue). The plot can be roughly divided into two halves: (a) planets with a mass of $< 100 M_\oplus$ and (b) planets with a mass of $> 100 M_\oplus$. In our simulations, most planets which started inside the ice line tend to have terrestrial Earth-like densities. These planets are $0.5\text{--}3 R_\oplus$ and $\lesssim 10 M_\oplus$. Planets starting around or outside the ice line generally accrete more volatile rich material and H/He gas. These planets have lower densities due to their larger sizes. Planet which started outside the ice line ($3\text{--}10$ au) show a broad diversity in their densities. As they accrete more gases, their density decreases further. These planets are roughly $2\text{--}10 R_\oplus$ and are characterised by masses that vary by four orders of magnitude. Planets more massive than $100 M_\oplus$ seem to lie on a single curve. Since the size of these planets remains the same ($\approx 1 R_J$ or $11 R_\oplus$), their densities increase linearly with their masses. Planets that started in the outer regions ($30\text{--}40$ au) cluster on the density-mass plane. These planets have low densities ($< 2\text{g cm}^{-3}$) and low masses ($\lesssim 1 M_\oplus$).

The density architecture for simulated systems in the Bern Model is shown in Fig. 9 (middle). An important relation between mass architecture and density architecture is seen. Some systems which are ordered (or anti-ordered) in mass are also ordered (or anti-ordered) in density, that is, these systems have large positive (or negative) $C_S(\rho)$. In other words, simulations suggest that planetary systems can also be ordered or anti-ordered in density. A system is ordered in density when the inner planets have small densities and the outer planets have larger densities – and vice-versa for density anti-ordered systems. Systems with mass architectures of similar and mixed are strongly clustered around $C_S(\rho) \approx 0$ and $C_V(\rho) < 1$. The inset shows that similar mass systems tend to have small $C_V(\rho)$, while mixed mass systems have larger $C_V(\rho)$. This implies that some systems that are similar (or mixed) in mass show some similarity (or mixture) in density. A system with a *similar* density architecture will host planets that have approximately similar densities. However, the region $C_S(\rho) \approx C_V(\rho) \approx 0$ is empty, indicating the absence of planetary systems where the density of planets (inside out) is approximately the same. While there are exceptions, overall,

⁸ A future study could investigate the boundaries for robust architecture identification, as in Eq. (3), but based on radius instead of mass. Such a classification is readily applicable since radius measurements tend to be uniformly available and are better agreed upon amongst several observers. Data-driven approaches such as machine learning could be useful in such an endeavour.

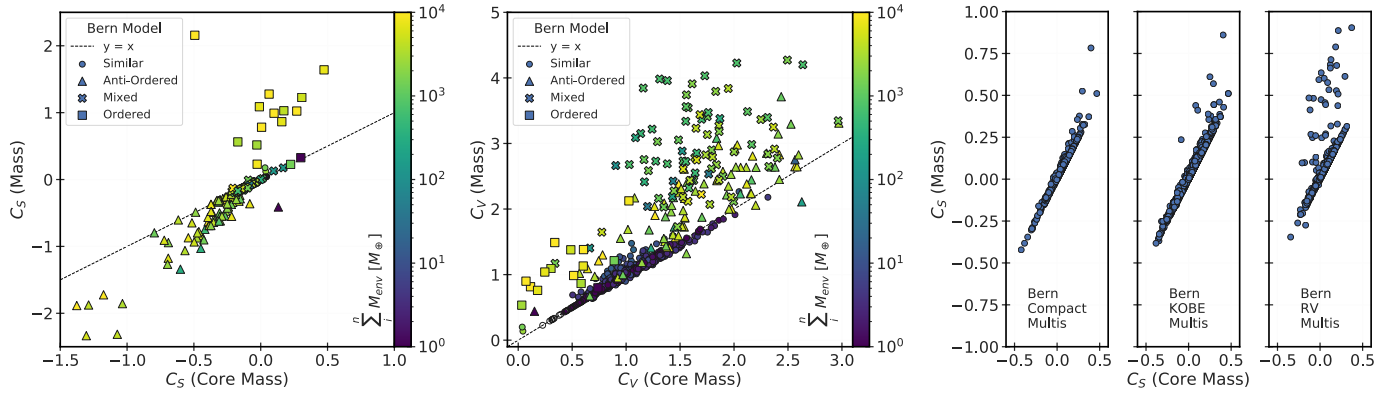


Fig. 10. Mass architecture as a function of core-mass architecture. Panels compare the mass architecture with the core-mass architecture via the coefficient of similarity (left) and coefficient of variation (middle). In the left panel, the points corresponding to similar systems are very tightly clustered on the $y = x$ line and are not visible due to over-plotting of points from other architectures. This signifies the core-mass architecture is very strongly correlated with the mass architecture for similar systems. The sum of mass in the envelope of each planet in a system is indicated in colour. The right panel plots the coefficient of similarity for masses and core masses for systems in the synthetically observed catalogues.

for many systems, the density architecture seems to follow their mass architecture.

This approximate link between the mass and density architecture stems from massive planets ($>100 M_{\oplus}$) whose densities increase with their mass (see Fig. 9 (left)). Systems which do not host any massive planet are mostly similar in their mass architecture and have $C_S(\rho) \approx 0$. The inset shows that the Aryabhata’s number increases as a system approaches the $C_S(\rho) \approx C_V(\rho) \approx 0$ region (see Paper II for the definition of Aryabhata’s number). If a system has more surviving planets that started from inside the ice line, then the densities of these planets will be more similar to each other. This means that the density architecture of a system shows some dependence on the starting location of a planet.

We also investigated if the relation between the mass and density architectures is observable. Figure 9 (right) shows the density architecture for systems from our synthetically observed catalogues. Also shown is the density architecture of some observed exoplanetary systems for which the mass and radius measurements were available. The density architecture of synthetically observed catalogues shows a trend which is quite unlike Fig. 9 (middle). There is an unexpectedly good agreement between the synthetically observed systems and the observed planetary systems. We attribute the peculiar shape of this plot to the difficulty of detecting distant planets. Transit and RV observations favour the detection of planets within ~ 1 au. Many close-in planets tend to have Earth-like densities, while planets farther out have lower densities (due to either their volatile rich or gaseous composition). Overall, this would lead to an observed density architecture where inner planets have higher densities and outer planets have lower densities. A situation such as this will be characterised by negative $C_S(\rho)$, which is readily seen from Fig. 9 (right).

In summary, many synthetic systems show a relationship between their mass architectures and their density architectures. Bern model systems that are ordered or anti-ordered in their mass also tend to be ordered or anti-ordered in their densities. The dispersion of planetary bulk densities in *similar* class systems is lower than mixed class systems. This relation seems to emerge from massive planets whose densities increases linearly with their masses (since they cannot grow their sizes any more). These relations can be considered as a prediction from this work. As future observations probe the outer parts of an exoplanetary

system, we may anticipate the discovery of several systems whose mass and density architectures are closely linked.

5.3. Core and envelope mass architecture

In this section, we show that (a) most simulated planetary systems inherit their architecture from the underlying core mass architecture; (b) the accretion of gases tends to accentuate the underlying core mass architecture, and (c) the observed mass architecture of a planetary system is a gateway to studying the core mass architecture of the system, since the two are strongly correlated. Exceptions to the first two statements tend to arise for those systems undergoing strong, multi-body dynamical effects such as planet-planet scattering.

The fraction of mass which is partitioned into a planet’s core and its envelope is governed by planetary formation physics. The end result is dictated by an interplay of several concurrent processes (see Emsenhuber et al. 2021a; Mordasini et al. 2012b, for discussion). In the core-accretion scenario, giant planets are formed when planetary cores can undergo run-away gas accretion (Pollack et al. 1996; Alibert et al. 2004, 2005). Proto-planets that have failed to trigger runaway gas accretion comprise a diverse group of planets: Earths, Super-Earths, mini-Neptunes, and Neptunes.

The bifurcation of a planet’s mass into its core and its envelope can probe the formation history. For example, in our simulations, most giant planets ($\gtrsim 1 M_J$) have about 1% of their mass in their cores and the rest is in their gassy envelope. On the other hand, low mass planets ($\lesssim 10 M_{\oplus}$) hardly accrete any gaseous envelope. However, the mass in a planet’s core and envelope is not an observable. Even for the solar system planets, internal structure models guide our knowledge of core and envelope masses (see Helled et al. 2020, for a review on Uranus and Neptune).

As giant planets dominated by their H/He envelopes are rare, we expect a strong correlation between the mass architecture (i.e. the arrangement and distribution of planetary masses) and the core-mass architecture (i.e. the arrangement and distribution of core-masses) to exist also at the system level. In Fig. 10, we show the coefficient of similarity and the coefficient of variation of planetary mass as a function of the coefficient of similarity and the coefficient of variation of core mass. The colour indicates the total mass of envelope accreted by all planets in a system.

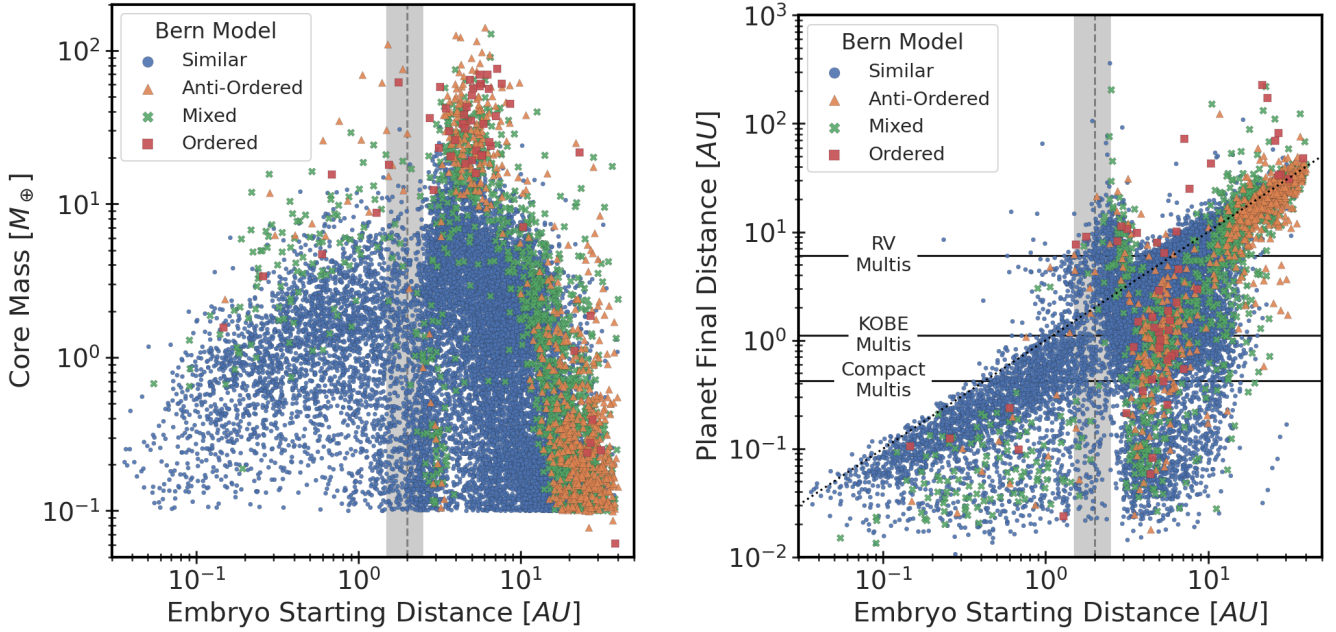


Fig. 11. Role of starting location. Plot shows the planetary core mass (left) and final distance (right) versus the starting distance. The marker style indicates the architecture of the system to which the planet belongs. The vertical grey shaded region indicates the evolving locations of the ice line (Burn et al. 2019). The dotted line in the right panel shows the $y = x$ line.

Comparing the coefficient of similarity for planetary masses and core masses (Fig. 10, left panel), we observe that a large fraction of systems (>90%) follow the $y = x$ line. This implies that for most planetary systems, the arrangement and distribution of core masses is imprinted on the mass architecture of the system. Systems which show large deviations from the $y = x$ line have generally accreted a large amount of gaseous envelope. This suggests that the formation of one or more giant planet is partly responsible for the deviations. We also observe another important feature. Planetary systems that are ordered in mass are also often ordered in their core-masses. Conversely, mass anti-ordered systems tend to be anti-ordered in their core masses as well. In addition, ordered systems are either on or above the $y = x$ line, whereas anti-ordered systems are either on or below this line. This suggests that the accretion of gases generally accentuates the underlying core mass architecture.

Considering the coefficient of variation for masses and core masses (Fig. 10, middle), we see that most of the planetary systems lie either on or above the $y = x$ line. The C_V value measures the amount of variation in a set of numbers. This suggests that the variation in total masses, for most systems, is either similar or larger than the variation in the core masses. This is understandable, since the amount of gas accreted by a planet shows a strong correlation with the mass of the planet’s core. However, there are a handful of systems where the variation in total mass is less than the variation in core masses. Systems that are similar in the mass architecture are strongly clustered over the $y = x$ line. This stems from the low amount of gas ($0\text{--}20M_{\oplus}$) accreted by planets in these systems. Figure 10 (middle) shows that mixed class systems, as opposed to similar systems, form a separate cluster. Physically, this difference is arising from the larger amount of gas ($50\text{--}5000M_{\oplus}$) accreted by planets in these systems.

Here, the question arises as to whether the strong correlation between mass architecture and core-mass architecture is observable. In Fig. 10 (right), we show $C_S(M)$ as function of $C_S(M_{\text{core}})$ for the three synthetically observed catalogues.

All three catalogues probe the inner regions of a planetary system. The figure shows that the correlation between mass architecture and core mass architecture is strong in all three catalogues. This suggests that the observed mass architecture of a planetary system can be used to study the underlying core-mass architecture of the system. This is potentially useful to distinguish among competing models of planet formation.

Role of embryo starting location. We have seen that the core mass architecture of a system strongly governs the overall architecture of the system. The arrangement of planets in a system also reflects the final distances of these planets. It is, therefore, instructive to understand some key aspects which shape these two important properties. The core mass and the final distance of a planet are strongly influenced by, among other effects, the distance at which an embryo starts in our simulations. Figure 11 shows the core mass (left) and the final distance (right) as a function of the starting distance. In the Bern model, lunar mass ($0.01M_{\oplus}$) protoplanetary embryos are initialised with a random starting location between the inner edge of the disk and 40 au. We also recall that failed embryos (objects with a total masses $<0.1M_{\oplus}$) are removed from our analysis.

Emsenhuber et al. (2021a); Burn et al. (2021) analysed the nature of planetary migration using migration maps. Both studies show the existence of so-called convergence zones. Within these zones, planets can migrate outwards. However, outside this zone inward migration is prevalent. The existence of such convergence zones suggests that there ought to be regions of planet over-densities; these are essentially regions where planets are radially ‘stuck.’ These studies attribute the presence of these zones to dust opacity transitions and disc structures, finding that these zones evolve with the disc. For a $0.01M_{\odot}$ disc, around a solar mass star at 1Myr, these zones are: (a) for low-intermediate mass planets ($\lesssim 1M_{\oplus}$) extending from disk inner edge to about 1au and (b) for intermediate mass planets ($1\text{--}10M_{\oplus}$) around 2–3 au.

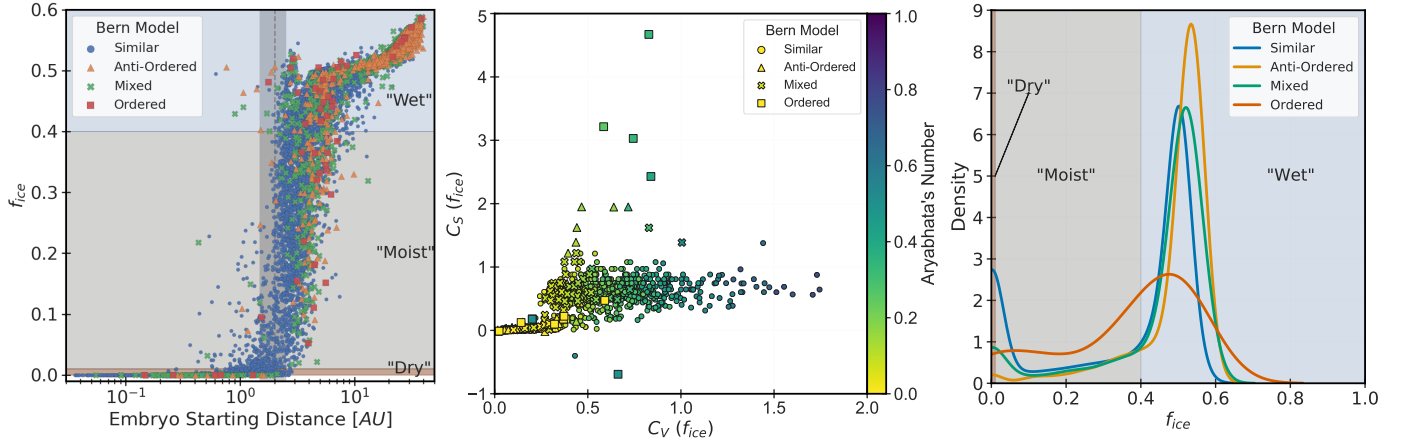


Fig. 12. Planetary core water-ice mass fraction. Left: core water mass fraction of a planet as a function of its starting location. The architecture of the system to which a planet belongs to is shown by marker characteristics. The vertical shaded regions shows the location of the ice line. Middle: water mass fraction architecture seen through the coefficient of similarity versus the coefficient of variation plot. The shape of the marker shows a system’s mass architecture, and the colour depicts its Aryabhata’s number (see Paper II for definition). Right: distribution of f_{ice} across architecture classes. Depending on f_{ice} , planets are labelled as ‘dry’, ‘moist’, or ‘wet’.

Figure 11 (left) shows that even for embryos that start at the same initial distance, the mass accreted by a planetary core can differ by two to three orders of magnitude. These differences arise from (a) varying solid disc masses; (b) competition for accretion in the feeding zone (Alibert et al. 2013); (c) dynamical state of solids in the disc resulting from planetesimal-planetesimal, planetesimal-protoplanet, planetesimal-gas disc interactions, and so on. Nevertheless, the starting distance seems to play a significant role in this scenario. The ice line seems to divide the parameter space into two regions: fewer planets inside the ice line have low mass cores ($\lesssim 1 M_{\oplus}$), while many planets outside the ice line have low-mass cores.

Inside the ice line, most planets have cores of $1\text{--}10 M_{\oplus}$. Planets that start very close to the star ($\lesssim 0.1$ au) are unable to accrete a lot of material owing to their small Hill spheres. This explains their small cores masses. Inside the ice line, planets belonging to systems of mixed, anti-ordered, and ordered architecture tend to have more massive cores than planets belonging to similar systems. Around the ice line, planets show a large variety of core masses ranging from $0.1 M_{\oplus}$ to $100 M_{\oplus}$. Outside the ice line we see the same trend as before: planets that are in similar systems, for the same starting location, usually have less massive cores than planets which belong to systems of other architectures.

The final distance of a planet depends on several factors such as: (a) migration type (type I or type II), (b) planet’s mass, (c) local disc properties, and (d) multi-body effects such as N -body scattering. The joint distribution of a planet’s final and starting locations shows an intriguing trend. Generally, for many planets, the final distance strongly correlates with their starting location. Orbital migration allows planets to move (mostly) inwards – positioning many planets below the $y = x$ line. N -body effects (such as planet-planet scattering or outward migration) may scatter some planets further away from their host star. These planets are located above the $y = x$ line. Curiously, many planets which end up farther away than their starting location were initialised around the ice line and are mostly low massive ($\lesssim 20 M_{\oplus}$). We attribute this over-density to the outward migration convergence zone around the ice line discussed above.

Another important finding is that planets inside the ice line in similar systems probably formed in situ. Figure 11 (right) shows that most planets, inside the ice line, which did not migrate

inwards are part of similar architecture systems. Conversely, most of the planets which have migrated inwards seem to belong to systems that have mixed, anti-ordered, and ordered architectures. Outside the ice line, many planets have migrated inwards. Most planets starting around 20 au (or more) accrete little mass in their cores and show little radial displacement (Hansen & Murray 2012; Chiang & Laughlin 2013). The properties of these embryos may draw some influence from our modelling choice as well. The N -body integrator in this model is used for 20 Myr. Longer integration times may allow some embryos to have more massive cores via giant impacts.

5.4. Core water-ice mass fraction architecture

Our model calculates the internal structure of a planetary core (for details see Emsenhuber et al. 2021a; Mordasini et al. 2012a). We solved 1D differential equations demanding mass conservation and hydrostatic equilibrium, with a modified polytrope equation serving as the equation of state (Seager et al. 2007). The chemical composition of each planetary core is also tracked. This is accomplished by tracking the chemical makeup of the accreted planetesimals and other colliding planets. The underlying chemical models have thirty-two refractory and eight volatile species (Thiabaud et al. 2014; Marboeuf et al. 2014a,b). These different chemical species are grouped into three different materials which make the planet’s core, in our model: (a) iron, (b) silicates, and (c) ice. All refractory species (except iron) make up the silicate mantle and all volatile species contribute to ice. Since H_2O constitutes 60% of all ice by mass, we label this latter component as water ice. The water mass fraction (f_{ice}) of each planetary core is computed.

We assume that inside the H_2O ice line, only refractory elements contribute to the solid phase of a planetesimal. Outside this evolving ice line, due to their condensation, volatile elements also contribute to the solid phase of a planetesimal. Figure 12 (left) shows the water mass fraction of a planet’s core as a function of its initial location. Most planets which start inside the ice line have little to no volatiles in their cores. A jump in f_{ice} is seen around the ice line. Outside the ice line, most planets have at-least 40% f_{ice} in their cores. This suggests that the history of

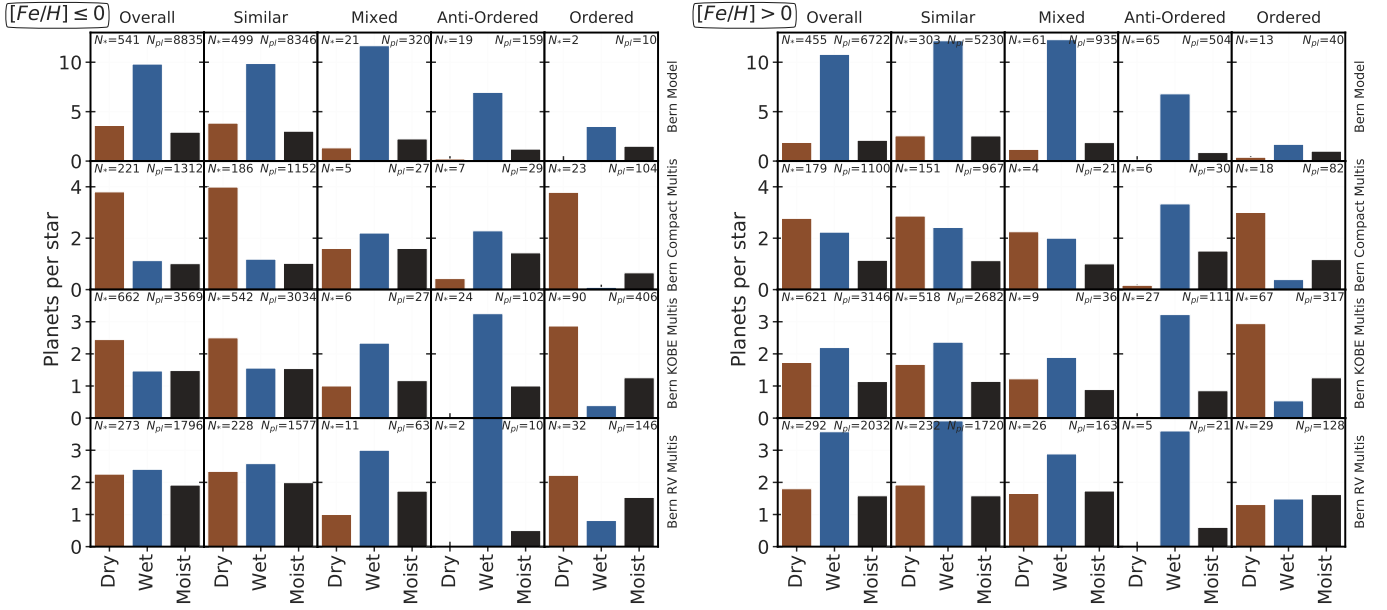


Fig. 13. Frequency of planets. This diagram shows the average planet per star for dry, wet, and moist planets in several catalogues (rows), across several architecture classes (columns), and around low (left) and high (right) metallicity stars. The planet per star is simply the total number of planets divided by the total number of stars, after appropriate filters for metallicity, catalogue, or architecture.

formation and evolution of a planet is imprinted on its water mass fraction.

We are interested in studying the ice mass fraction architecture of a planetary system. However, we cannot directly apply our framework (Eqs. (1) and (2)) because the water mass fraction is a quantity that admits 0 as a value. While this can lead to ill-defined numbers, this issue has a simple remedy. For quantities that can be 0, we propose the following modification to Eq. (1):

$$C_S(q) = \lim_{\epsilon \rightarrow 0} \frac{1}{n-1} \sum_{i=1}^{i=n-1} \left(\log \frac{q_{i+1} + \epsilon}{q_i + \epsilon} \right). \quad (4)$$

Numerically, we calculated the coefficient of similarity with $\epsilon = 10^{-10}$. We verified this step by calculating the coefficient of similarity for quantities which do not admit zero (such as masses). In a bootstrapped numerical experiment of 10,000 trials, the coefficient of similarity for mass was calculated using both Eq. (1) and (4). The relative difference between the two outcomes ranged between 10^{-12} and 10^{-10} .

The ice mass fraction architecture of Bern Model systems is shown in Fig. 12 (middle). A prominent feature from this figure is that most systems have $C_S(f_{ice})$ either close to 0 or positive. A system with $C_S(f_{ice}) \approx 0$ and low $C_V(f_{ice})$ will be composed of planets whose core water mass fraction is similar to one another. A system with positive $C_S(f_{ice})$ will be composed of planets whose core water mass fraction increases inside out. Figure 11 (right) tells us that many planets that started outside the ice line, and are water rich have not suffered any major radial displacement. Thus, a positive $C_S(f_{ice})$ should be a default scenario for most planetary systems. About 74% systems in the Bern model have $C_S(f_{ice}) > 0.1$. Almost 97% of systems have $C_S(f_{ice}) > 0$. We propose the ‘Aryabhata formation scenario’ to explain the ‘non-default’ systems. This scenario and the related quantity ‘Aryabhata’s Number’ are described in Paper II.

5.5. Frequency of dry, moist, and wet planets

We are interested in exploring the link between the water mass fraction architecture and the mass architecture of a system. To this end, we divide planets into three categories based on their water mass fraction. A planet is called ‘dry’ if $f_{ice} \leq 1\%$, ‘moist’ if $f_{ice} \in (1, 40]\%$, and ‘wet’ if $f_{ice} > 40\%$. These labels serve to simplify our analysis and allows us to see general trends between system architecture and planetary composition. The distribution of water mass fraction across systems of different architecture classes is shown in Fig. 12 (right). While all three planet classes are present in all four architecture classes, there are some observable trends.

Figure 12 (right) shows that similar architectures host many of the dry planets produced in the Bern model and anti-ordered architectures are mostly composed of wet planets. This tells us that many of the planets that start inside the ice line become part of similar architecture systems. Conversely, systems with anti-ordered architecture are mostly composed of planets that started outside the ice line. Mixed architecture systems are generally composed of more planets that started outside the ice line than inside, as compared to similar architecture systems. Moist planets are present in all architecture classes. We quantify the frequency of dry, moist, and wet planets as a function of mass architecture class (similar, mixed, ordered, or anti-ordered), metallicity (low or high), and source catalogues (Bern model, Bern Compact Multis, Bern KOBE Multis, and Bern RV Multis). Figure 13 shows the planets per star (i.e. the number of each planet type divided by the number of stars) across these forty sub-categories.

Overall, compared to synthetically observed catalogues, Bern model simulations demonstrate more wet planets. This is understandable since we are looking at the entire underlying population, which includes planets from the outer parts of these systems. Likewise, synthetically observed catalogues tend to have more dry planets. Systems around low-metallicity stars (regardless of the catalogue) generally tend to have a

higher frequency of dry planets as opposed to systems around high-metallicity stars. The frequency of wet planets shows a noticeable increase for systems around high-metallicity stars. Amongst the different catalogues, Bern Compact Multis have the highest frequency of dry planets, followed by Bern KOBE Multis, and Bern RV Multis. Low-metallicity environments have a slightly higher average planet per star ($8835/541 \approx 16.3$) than high-metallicity environments ($6722/455 \approx 14.8$).

Similar systems. Systems in the underlying Bern model that are characterised by a similar architecture tend to have many wet planets (~ 10 per star) and few dry or moist planets ($\sim 3\text{--}4$ per star). However, synthetically observed catalogues seem to have a bias against the discovery of many wet planets. For the similar class of compact multi-planetary systems, dry planets are more common around a low-metallicity star. However, for a high-metallicity star, the frequency of dry and wet planets is roughly the same. For transiting systems, in the Bern KOBE Multis, low-metallicity environments favour more dry planets and equal proportions of wet and moist planets. Conversely, in high-metallicity environments, wet planets occur more frequently than dry or moist planets. For RV systems, the frequency of each planet class is approximately the same in a low-metallicity environment. High-metallicity environments almost double the frequency of wet planets. The average planet per star is similar around both low metallic (≈ 16.8) and high metallic environments (≈ 17.3).

Mixed systems. Mixed class systems generally have many wet planets. It is only for compact systems around high-metallicity stars, the frequency of dry planets is higher than wet planets. In all other cases, the frequency of wet planets is greater than the frequency for dry or moist planets. The average planet per star is similar around both low-metallicity (≈ 15.2) and high metallicity environments (≈ 15.3).

Anti-Ordered systems. Systems with anti-ordered architecture have a distinct core water mass fraction architecture. These systems are rich in wet planets. In fact, about 80% of these systems follow the Aryabhata formation scenario described in Paper II. Compact anti-ordered systems may have some dry planets. For transit and RV surveys, the frequency of dry planets is zero in our simulations. The total number of planets per star in anti-ordered systems is slightly higher around low metallicity stars ($159/19 \approx 8.4$), as compared to high metallicity stars ($504/65 \approx 7.8$). In the future, if an anti-ordered architecture planetary system is to be discovered, it would be interesting to study its core water mass fraction architecture as well. The current work suggests that the Aryabhata's number for these systems should be close to 0 and, irrespective of the detection technique, the system should be expected to have many wet planets (see Paper II); this is one of the main predictions arising from this work.

Ordered systems. Juxtaposed directly to the anti-ordered systems, ordered systems in synthetically observed catalogues tend to be rich in dry planets. These systems are distinct not only because of their frequent dry planets, but also due to a low frequency of wet planets. For all synthetic catalogues, moist planets occur more frequently than wet planets, which is a unique distinguishing feature for these systems. For the Bern model, these systems have low average planets per star: 5 around low-metallicity stars and 3.1 around high-metallicity stars.

In summary, we note some salient features of these system architectures. Generally, wet planets survive more frequently

around high-metallicity stars. One detection technique that favours the discovery of close-in planets also favours the detection of dry planets. The comparative frequency of planet (dry, wet, or moist) per star seems to be intimately connected with the mass architecture of the system. Similar and mixed systems can host lots of dry or wet planets, depending on the metallicity of the systems and detection technique. Anti-ordered systems, forming prominently via the Aryabhata formation scenario, are rich in wet planets. Ordered systems, in simulated observations, are rich in dry planets and have more moist planets than wet planets. The physical connection between the average planet per star and the star's metallicity is sensitive to the formation pathways that a system undergoes.

6. Habitability as a function of system architecture

In this paper thus far, we have described a new framework for studying the architecture of planetary systems (Sect. 3), the characteristics of the four classes of system architecture (Sect. 4), and the relation between the mass architecture of a system and its internal structure and composition architecture (Sect. 5). In this section, we speculate on the idea of studying habitability as a function of system-level architecture.

Mankind has pondered the existence of other biotic life-forms beyond Earth, as well as outside our own Solar System. Our current understanding of habitability stems from and is focused at an individual planetary level. We consider whether habitability could be correlated with other properties of a planetary system, namely, whether habitability could be a system-level phenomenon. In this section, we speculate on the role of planetary-system level information on the existence of habitable worlds in such systems. The framework we present here for studying the system-level architecture of a planetary system brings to light several novel questions, probing the dependence of habitability and occurrence of habitable worlds (and related concepts) on the architecture of a said system. For example, we wonder how the occurrence rate of habitable planets in the galaxy depends on the occurrence of the four architecture classes.

In this section, we address this question on three levels: system, planet, and planet ratio. We use the concept of empirical Habitable zone (EHZ) planets from Quanz et al. (2022); Kopparapu et al. (2014). Planets with masses between $[0.1, 5]M_{\oplus}$ and stellar insolation within $[1.776, 0.32]S_{\oplus}$ are considered to be inside the EHZ. The stellar flux limits correspond to 'recent Venus' and 'early Mars' scenarios and include the luminosity evolution for a $1 M_{\odot}$ Solar-twin. At the system level, we note the frequency of systems of a particular architecture to host at least one planet in the EHZ. At the planet level, we count the frequency of planets in the EHZ across each system architecture class. At the planet ratio level, we show the fraction of all EHZ planets across their architecture class. Figure 14 shows the frequency of EHZ planets, at all three levels, as a function of their system architecture for both synthetic and observed exoplanetary systems.

Out of all synthetic systems with a similar class architecture, $\approx 77\%$ host at least one EHZ planet. This is remarkably higher than any other architecture class. $\approx 10\%$ of systems with mixed architecture host at least one EHZ planet. The frequency drops to $\approx 1\%$ for anti-ordered architecture systems and $\approx 0\%$ for ordered systems. One way to interpret these numbers could be to look at the multiplicity distribution across each architecture class in Fig. 7. The frequency of at least one EHZ planets across

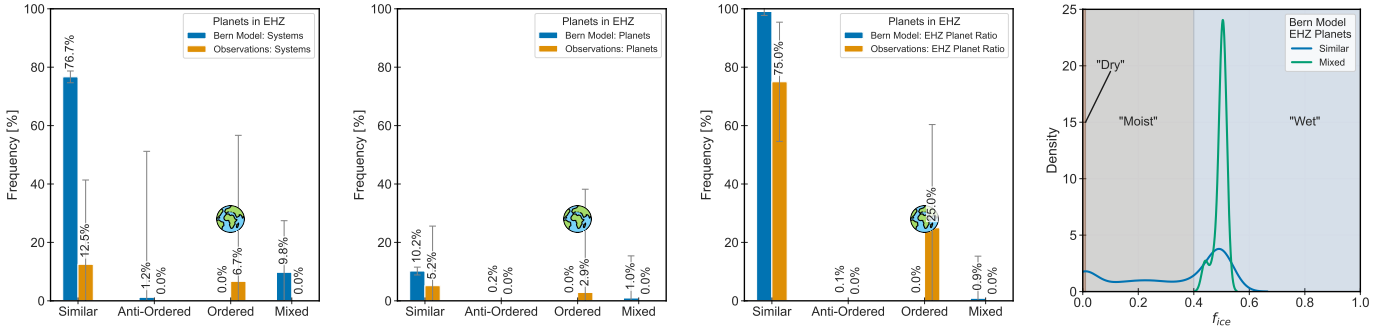


Fig. 14. Planets inside the empirical habitable zone (EHZ). The left-most plot shows the frequency of planetary systems, of a given architecture class, which host at least one planet inside the EHZ. The central-left plot shows the fraction of planets inside a given architecture class which are in the EHZ. The central-right plot shows the fraction of all EHZ planets within a given architecture class. The rightmost plot shows the distribution of f_{ice} for EHZ planets across the architecture classes. The cartoon sketch of Earth emphasises that the only known life-harboring planet resides in an ordered system. The length of error bars visualises the total number of systems or planets in respective bin as: $100 / \sqrt{\text{bin counts}}$. The lengths of the error bars represents the number of planetary systems (left-most panel) and the number of planets (two middle panels) which are inside the bin. Large error bars in the leftmost panel, for example for anti-ordered architecture emerges from their low count (see Fig. 5). The Gaussian kernel is estimated using Scott’s rule (Scott 2015).

architecture class seems to follow the multiplicity trends. Similar and mixed architectures have comparably high number of planets. The distribution of the Aryabhata’s number shows that similar systems usually have higher Aryabhata’s number than mixed systems, implying that similar systems tend to host more planets which started from inside the ice line (see Paper II for Aryabhata’s number). This may account for the large frequency of similar systems which host at least one EHZ planet. The multiplicity distribution shows that anti-ordered systems often host less planets than similar and mixed class systems, while ordered systems have the lowest multiplicities. We see in Sect. 4.2 that the similar class architecture is perhaps the most common architecture for planetary systems in our galaxy. These results from the Bern model simulations suggest that observation campaigns to detect habitable planets will find more EHZ planets in similar class architectures.

For the observed multi-planetary systems in our catalogue, about $\approx 13\%$ of similar class systems have at least one EHZ planet. About 7% of ordered class exoplanetary systems in our catalogue host at least one EHZ planet. In the mixed class observed systems in our catalogue, none of them have EHZ planets and there are no known anti-ordered class systems in our catalogue. These frequencies are quite different from their theoretical counterparts. While the lack of a complete and reliable observations catalogue may explain the discrepancy for similar class systems – it does not completely explain the discrepancy for ordered systems. Our own planet resides in the ordered class system of the Solar System, which is not supposed to be influenced by issues such as completeness or detection biases. This reflects the inability of Bern models to simulate a Solar System analogue – pointing to a gap in our understanding of the physics that goes into planetary formation and evolution. In addition, many observed ordered class systems may have a different architecture when more planets in these systems are detected.

At the planet level in our simulations, out of all synthetic planets that exist in similar class systems, about 10% are inside the EHZ. This frequency is, again, remarkably higher for any other architecture class. About 1% of all simulated planets in a mixed system are inside the EHZ. Close to 0% of all planets in anti-ordered and ordered class architectures are inside the EHZ. From our observational catalogue, while 5% of observed exoplanets in similar class systems are inside the EHZ. About 3% of observed exoplanets in ordered class systems are inside the EHZ.

The planet ratio level shows the fraction of all EHZ planet that belong to a particular architecture class. In the Bern model, we see that out of all EHZ planets, about 99% are in the similar class. The share of EHZ planets by other architecture classes is negligible. Amongst the observations, three-quarters of EHZ planets are in similar class and the remaining are in ordered class. The observations and theory are quite misaligned in this scenario. We attribute this discrepancy to the absence of a complete and reliable catalogue of observations.

Our observations catalogue has only 41 multi-planetary systems, of which only four host planets inside the EHZ. These systems are Trappist-1 (three planets in EHZ), GJ 667 C (two planets in EHZ), Solar System (two planets in EHZ), and Tau Ceti (one planet in EHZ). The occurrence of architecture classes and the frequency with which they host EHZ planets might be better constrained with future observations. This may allow us to have a better estimate of the occurrence rate of EHZ planets as a function of architecture class.

Simulations suggest that ordered architecture is a rare outcome of planet formation (about 1.5% of systems out of 1000 were deemed to be ordered) and yet, we live in an ordered system. These two statements can shed new light on the rarity of life in the galaxy. We foresee that the famous Drake equation may be suitably modified to take into account the occurrence rate of different architectures and thereby set more optimal constraints on η_{\oplus} (Sarkar 2022).

Since water plays a fundamental role for life forms on Earth, it is interesting to probe the core water-ice fraction for the EHZ planets. Figure 14 also shows the f_{ice} distribution for EHZ planets in the Bern model. As we see before, most of the EHZ planets are in the similar class and $\approx 1\%$ of EHZ planets are in the mixed class. EHZ planets in similar systems are ‘dry’, ‘moist’, and ‘wet’. In stark contrast, EHZ planets in mixed class are only ‘wet’ planets. We hope these results may be useful in guiding future missions in finding EHZ planets that have the potential to harbour life.

7. Summary, conclusions, and future work

In this paper, we introduce and explore a new framework for studying the architecture of planetary systems. Our new framework allows us to study, quantify, classify, the global architecture

of an entire planetary system at the system-level; and compare the architecture of one planetary system with another. In Sect. 3, we detailed the new architecture framework and presented an in-depth discussion comparing our framework with other works in the literature. We present the coefficient of similarity and the coefficient of variation as two quantities that quantify our conceptual ideas. Our framework gives rise to a new parameter space (the C_S versus C_V plane) in which individual planetary systems can be compared with one another. Throughout this paper, we applied this framework to study the distribution and arrangement of several planetary quantities within a planetary system, thereby understanding the system architecture for that quantity. In this manner, we studied the mass architecture, the radius architecture, the core mass architecture, the core water mass fraction architecture, and the density architecture of synthetic and observed planetary systems.

To study some consequences of this framework, we applied our method to several catalogues of planetary systems (introduced in Sect. 2). We curated, especially for the purposes of this study, a catalogue of observed multi-planetary systems that have four or more planets and include mass measurements for at least four planets. For engendering further studies, additional stellar and planetary properties were collected and presented in Table 1. We also used synthetic planetary systems simulated via the Bern model. To facilitate a comparison of theory with observations, we prepared three synthetic observed catalogues by applying the detection biases on the simulated planetary systems. This led to the Bern RV Multis, Bern KOBE Multis, and the Bern Compact Multis catalogue. We note that there are caveats present in the datasets we used. The model-dependent results we present here may be improved upon in future studies using better theoretical models and a more complete observational catalogues (e.g. from PLATO).

Summary of architecture framework:

1. The architecture framework is model-independent and therefore does not suffer from any caveats emerging from planet formation theory or observations.
2. The same architecture framework can be used to study the multi-faceted aspects of planetary system architecture. When the framework is applied to study planetary masses, the framework informs us of the mass architecture of the system, namely, the arrangement and distribution of masses in the planetary system. In this way, we can use this framework to study the mass architecture, radii architecture, eccentricity architecture, and so on for the same system. In this series of work, we identified the architecture of a system with its bulk mass architecture.
3. Planetary system architecture can be one of four classes that are derived from our framework: similar, mixed, ordered, and anti-ordered.
4. A planetary system's architecture is of similar class when the masses of all the planets within such a system are similar to each other. This architecture class corresponds to the 'peas in a pod' architecture trend reported in the literature.
5. The architecture class of a planetary system is ordered (or anti-ordered) when the planetary masses in such systems tend to increase or decrease from inside-out.
6. Planetary systems of mixed class architecture host planets whose masses show broad increasing and decreasing variations.

Our key model-dependent findings are as follows:

1. Frequency of architecture class: systems with similar bulk mass architecture are the most common outcome of

simulations, followed by the other three architecture classes. Our model suggests that similar architecture should be the most common exoplanetary system architecture in our Galaxy and beyond. This explains why radius similarity in exoplanets was already detected from the first four months of *Kepler* data (Lissauer et al. 2011).

2. Distance bi-modality: we found hints of a bi-modality in the exoplanetary distance distribution arising from the two different modes of orbital migrations. This bi-modality is readily visible (see Fig. 7) for similar and mixed mass architecture exoplanetary systems observed via RV.
3. Core mass architectures: we found that for most systems, the bulk mass architecture is inherited from the core mass architecture. In addition, the accretion of gases tends to highlight the underlying core mass architecture by amplifying it. In this way, the observed mass architecture of a system could serve as a gateway for studying the distribution and arrangement of the planetary core masses, which tends to be simpler for theoretical modelling.
4. In situ formation: we found that most planets belonging to the similar bulk mass architecture class form in situ inside the ice line. In contrast, planets inside the ice line belonging to mixed, anti-ordered, and ordered show large inward migrations.
5. Core water-ice mass fraction architectures: synthetic planetary systems were found to have two scenarios for their core water mass fraction architecture. The default scenario consists of relatively more dry planets in the inner parts of a system and more wet planets in the outer parts of the system. This is probably a direct consequence of the starting location of planets: planets starting inside (or outside) the ice line tend to be dry (or wet). About one-fifths of simulated systems do not follow the default scenario described above. We propose the 'Aryabhata formation scenario' to explain their core-water mass fraction architecture (see Paper II).
6. Linking architecture and internal composition: we found that wet planets are more likely to survive around high-metallicity stars. Among other predictions, we showed that anti-ordered observed systems should be rich in wet worlds, while ordered observed systems are expected to have many dry planets (based on the core-accretion planet formation paradigm).
7. Density architectures: synthetic systems that are ordered (or anti-ordered) in mass tend to also be ordered (or anti-ordered) in their bulk densities. Some mass similar systems may also have low dispersion in their planetary bulk densities. The density architecture is sensitive to the Aryabhata's number (i.e. the starting location of various surviving planets; see Paper II). The density architecture of observed systems is in good agreement with the density architecture of synthetically observed simulated systems. Detection biases seem to favour the discovery of planetary systems where the densities show anti-ordering, mixing, or similarity.
8. Radius architectures: the radius architecture of most planetary systems closely follows their mass architecture. Therefore, most mass similar systems also show similarity in radius (also for mass mixed, ordered, or anti-ordered systems). However, this is not always true. Future studies can calibrate a classification scheme based on planetary radii.
9. Habitability as a system-level phenomenon: we reflected on the prospect of studying habitability as a function of system-level properties such as system architecture. Similar architecture systems represent an excellent observation

target for finding life outside the solar systems because these systems tend to host many more planets inside the empirical habitable zone than other architecture classes.

10. The current version of the Bern model seems to have difficulty in producing planets inside the EHZ of an ordered architecture system. Nevertheless, more data is required to conclude whether the existence of Earth, an inhabited planet in an ordered system, is an exception or whether there are additional gaps in our understanding of planet formation.

This paper is the first in a series. The current work presents a new testing ground, the architecture space, for theoretical models and for comparing observations with theory. We can now constrain our understanding of planet formation not only on the level of an individual planet – but at the global systemic level. This is a multi-faceted approach, since the system architecture of several quantities can now be uniformly assessed and compared with observations. In our next paper (Paper II), we show another important aspect emerging from this architecture framework which asserts that systems with comparable architecture often have the same formation pathways. We present ideas to further the nature versus nurture debate around planet formation. While similar architectures are usually a product of their starting conditions, stochastic multi-body effects are responsible for shaping the other three architecture classes. This work leads to several future studies which will be presented in other papers in this series. Davoult et al. (in prep.) explore how the present architecture framework can be employed for an efficient usage of telescope time to hunt for habitable worlds. Other possible explorations that emerge from this work include: (a) a data-driven approach to classifying planetary architecture based on radii and (b) a suitable modification to Drake’s equation that accounts for the empirical occurrence rate of system architectures.

Acknowledgements. The authors thank the anonymous referee for their careful reading, constructive suggestions, and insightful questions, which has allowed the quality of this manuscript to be improved. This work has been carried out within the framework of the National Centre of Competence in Research PlanetS supported by the Swiss National Science Foundation under grants 51NF40_182901 and 51NF40_205606. The authors acknowledge the financial support of the SNSF. Data: the synthetic planetary populations (NGPPS) used in this work are available online at <http://dace.unige.ch> under section “Formation & evolution”. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program: <https://exoplanetarchive.ipac.caltech.edu> (DOI: 10.26133/NEA6). The artwork used to depict Earth in Fig. 14 is taken from flaticon.com. Software: KOBE (Mishra et al. 2021; Mishra 2021), Python (Van Rossum & Drake 2009), NumPy (Oliphant 2006), SciPy (Virtanen et al. 2020), Seaborn (Waskom & the seaborn development team 2020), Pandas (pandas development team 2020), Matplotlib (Hunter 2007).

References

- Abdi, H. 2010, in *Encyclopedia of Research Design*, ed. N. Salkind (Thousand Oaks: SAGE Publications, Inc.)
- Adams, F. C. 2019, *MNRAS*, **488**, 1446
- Adams, F. C., Batygin, K., Bloch, A. M., & Laughlin, G. 2020, *MNRAS*, **493**, 5520
- Agol, E., Dorn, C., Grimm, S. L., et al. 2021, *Planet. Sci. J.*, **2**, 1
- Alibert, Y. 2019, *A&A*, **624**, A45
- Alibert, Y., Mordasini, C., & Benz, W. 2004, *A&A*, **417**, L25
- Alibert, Y., Mordasini, C., Benz, W., & Winisdoerffer, C. 2005, *A&A*, **434**, 343
- Alibert, Y., Mordasini, C., & Benz, W. 2011, *A&A*, **526**, A63
- Alibert, Y., Carron, F., Fortier, A., et al. 2013, *A&A*, **558**, A109
- Anglada-Escudé, G., Tuomi, M., Gerlach, E., et al. 2013, *A&A*, **556**, A126
- Archinal, B. A., Acton, C. H., A’Hearn, M. F., et al. 2018, *Celest. Mech. Dyn. Astron.*, **130**, 22
- Astudillo-Defru, N., Forveille, T., Bonfils, X., et al. 2017, *A&A*, **602**, A88
- Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, *A&A*, **577**, A42
- Bashi, D., & Zucker, S. 2021, *A&A*, **651**, A61
- Benatti, S., Damasso, M., Desidera, S., et al. 2020, *A&A*, **639**, A50
- Benz, W., Ida, S., Alibert, Y., Lin, D., & Mordasini, C. 2014, in *Protostars and Planets VI*, eds. H. Beuther, R. Klessen, C. Dullemond, & T. Henning (Tucson: University of Arizona), 691
- Berger, T. A., Huber, D., van Saders, J. L., et al. 2020, *AJ*, **159**, 280
- Boley, A. C., Hayfield, T., Mayer, L., & Durisen, R. H. 2010, *Icarus*, **207**, 509
- Bonfanti, A., & Gillon, M. 2020, *A&A*, **635**, A6
- Bonomo, A. S., Zeng, L., Damasso, C., et al. 2019, *Nat. Astron.*, **3**, 416
- Borucki, W. J., Agol, E., Fressin, F., et al. 2013, *Science*, **340**, 587
- Bouchy, F., Udry, S., Mayor, M., et al. 2005, *A&A*, **444**, L15
- Bourrier, V., Dumusque, X., Dorn, C., et al. 2018, *A&A*, **619**, A1
- Bryson, S., Kunimoto, M., Koppappu, R. K., et al. 2021, *AJ*, **161**, 36
- Buchhave, L. A., Dressing, C. D., Dumusque, X., et al. 2016, *AJ*, **152**, 160
- Burgasser, A. J., & Mamajek, E. E. 2017, *ApJ*, **845**, 110
- Burn, R., Marboeuf, U., Alibert, Y., & Benz, W. 2019, *A&A*, **629**, A64
- Burn, R., Schlecker, M., Mordasini, C., et al. 2021, *A&A*, **656**, A72
- Charbonneau, D., Berta, Z. K., Irwin, J., et al. 2009, *Nature*, **462**, 891
- Chevance, M., Diederik Kruijssen, J. M., & Longmore, S. N. 2021, *ApJ*, **910**, L19
- Chiang, E., & Laughlin, G. 2013, *MNRAS*, **431**, 3444
- Ciardi, D. R., Fabrycky, D. C., Ford, E. B., et al. 2013, *ApJ*, **763**, 41
- Coleman, G. A., & Nelson, R. P. 2014, *MNRAS*, **445**, 479
- Dalirsefat, S. B., da Silva Meyer, A., & Mirhoseini, S. Z. 2009, *J. Insect Sci.*, **9**, 71
- Díaz, R. F., Ségransan, D., Udry, S., et al. 2016, *A&A*, **585**, A134
- Dittkrist, K. M., Mordasini, C., Klahr, H., Alibert, Y., & Henning, T. 2014, *A&A*, **567**, A121
- Drazkowska, J., Bitsch, B., Lambrechts, M., et al. 2022, ArXiv e-prints [arXiv:2203.09759]
- Emsenhuber, A., Mordasini, C., Burn, R., et al. 2021a, *A&A*, **656**, A69
- Emsenhuber, A., Mordasini, C., Burn, R., et al. 2021b, *A&A*, **656**, A70
- Fortier, A., Alibert, Y., Carron, F., Benz, W., & Dittkrist, K.-M. 2013, *A&A*, **549**, A44
- Fressin, F., Torres, G., Rowe, J. F., et al. 2011, *Nature*, **482**, 195
- Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, *AJ*, **154**, 109
- Fulton, B. J., Rosenthal, L. J., Hirsch, L. A., et al. 2021, *ApJS*, **255**, 14
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Gáspár, A., Rieke, G. H., & Ballering, N. 2016, *ApJ*, **826**, 171
- Gilbert, G. J., & Fabrycky, D. C. 2020, *AJ*, **159**, 281
- Gillon, M., Triard, A. H. M. J., Demory, B.-O., et al. 2017, *Nature*, **542**, 456
- Gower, J. C. 1971, *Biometrics*, **27**, 857
- Goździewski, K., Maciejewski, A. J., & Migaszewski, C. 2007, *ApJ*, **657**, 546
- Gravity Collaboration (Lacour, S., et al.) 2019, *A&A*, **623**, L11
- Grimm, S. L., Demory, B.-O., Gillon, M., et al. 2018, *A&A*, **613**, A68
- Hadden, S., & Lithwick, Y. 2017, *AJ*, **154**, 5
- Hansen, B. M. S., & Murray, N. 2012, *ApJ*, **751**, 158
- Hara, N. C., Bouchy, F., Stalport, M., et al. 2020, *A&A*, **636**, L6
- Hartigan, J. A., & Hartigan, P. M. 1985, *Ann. Stat.*, **13**, 70
- He, M. Y., Ford, E. B., & Ragozzine, D. 2019, *MNRAS*, **490**, 4575
- He, M. Y., Ford, E. B., & Ragozzine, D. 2021, *AJ*, **161**, 16
- Hébrard, G., Arnold, L., Forveille, T., et al. 2016, *A&A*, **588**, A145
- Helffrich, G. 2017, *Prog. Earth Planet. Sci.*, **4**, 24
- Helled, R., Nettelmann, N., & Guillot, T. 2020, *Space Sci. Rev.*, **216**, 1
- Hsu, D. C., Ford, E. B., Ragozzine, D., & Ashby, K. 2019, *AJ*, **158**, 109
- Hueso, R., & Guillot, T. 2005, *A&A*, **442**, 703
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Jacobson, R. A. 2009, *AJ*, **137**, 4322
- Jacobson, R. A. 2014, *AJ*, **148**, 76
- Jacobson, R. A., Antreasian, P. G., Bordi, J. J., et al. 2006, *AJ*, **132**, 2520
- Jin, L., & Li, M. 2014, *ApJ*, **783**, 37
- Jin, S., Mordasini, C., Parmentier, V., et al. 2014, *ApJ*, **795**, 65
- Kalas, P., Graham, J. R., Chiang, E., et al. 2008, *Science*, **322**, 1345
- Kane, S. R., & Gelino, D. M. 2014, *ApJ*, **792**, 111
- Katsnelson, J., & Kotz, S. 1957, *Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B*, **8**, 103
- Kipping, D. 2018, *MNRAS*, **473**, 784
- Kokubo, E., & Ida, S. 1998, *Icarus*, **131**, 171
- Kokubo, E., & Ida, S. 2002, *ApJ*, **581**, 666
- Konopacky, Q. M., & Barman, T. S. 2018, in *Handbook of Exoplanets*, eds. H. J. Deeg, & J. A. Belmonte (Berlin: Springer), 36
- Koppappu, R. K., Ramirez, R. M., SchottelKotte, J., et al. 2014, *ApJ*, **787**, L29
- Koppappu, R. K., Hébrard, E., Belikov, R., et al. 2018, *ApJ*, **856**, 122
- Kratzer, K. M., Murray-Clay, R. A., & Youdin, A. N. 2010, *ApJ*, **710**, 1375
- Lacedelli, G., Malavolta, L., Borsato, L., et al. 2021, *MNRAS*, **501**, 4148
- Leleu, A., Alibert, Y., Hara, N. C., et al. 2021, *A&A*, **649**, A26
- Lillo-Box, J., Faria, J. P., Suárez Mascareño, A., et al. 2021, *A&A*, **654**, A60
- Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., et al. 2011, *ApJS*, **197**, 8

- Lissauer, J. J., Marcy, G. W., Rowe, J. F., et al. 2012, *ApJ*, **750**, 112
- Lissauer, J. J., Jontof-Hutter, D., Rowe, J. F., et al. 2013, *ApJ*, **770**, 131
- Lopez, E. D., & Fortney, J. J. 2014, *ApJ*, **792**, 1
- Lopez, T. A., Barros, S. C. C., Santerne, A., et al. 2019, *A&A*, **631**, A90
- Lovis, C., Ségransan, D., Mayor, M., et al. 2011, *A&A*, **528**, A112
- Lynden-Bell, D., & Pringle, J. E. 1974, *MNRAS*, **168**, 603
- MacDonald, M. G., Ragozzine, D., Fabrycky, D. C., et al. 2016, *AJ*, **152**, 105
- Marboeuf, U., Thiabaud, A., Alibert, Y., Cabral, N., & Benz, W. 2014a, *A&A*, **570**, A36
- Marboeuf, U., Thiabaud, A., Alibert, Y., Cabral, N., & Benz, W. 2014b, *A&A*, **570**, A35
- Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, *ApJS*, **210**, 20
- Marois, C., Macintosh, B., Barman, T., et al. 2008, *Science*, **322**, 1348
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, *The Messenger*, **114**, 20
- Mayor, M., Marmier, M., Lovis, C., et al. 2011, ArXiv e-prints [arXiv:1109.2497]
- Millholland, S. C., & Winn, J. N. 2021, *ApJ*, **920**, L34
- Millholland, S., Wang, S. C., & Laughlin, G. 2017, *ApJ*, **849**, L33
- Millholland, S. C., Laughlin, G., Teske, J., et al. 2018, *AJ*, **155**, 106
- Mills, S. M., Fabrycky, D. C., Migaszewski, C., et al. 2016, *Nature*, **533**, 509
- Mills, S. M., Howard, A. W., Weiss, L. M., et al. 2019, *AJ*, **157**, 145
- Mishra, L. 2021, Astrophysics Source Code Library, [record ascl:2106.001]
- Mishra, L., Alibert, Y., & Udry, S. 2019, in EPSC-DPS Joint Meeting 2019, held 15-20 September 2019 in Geneva, Switzerland, EPSC-DPS2019-1616, 2019
- Mishra, L., Alibert, Y., Leleu, A., et al. 2021, *A&A*, **656**, A74
- Mishra, L., Alibert, Y., Udry, S., & Mordasini, C. 2023, *A&A*, **670**, A69 (Paper II)
- Mordasini, C. 2018, in *Handbook of Exoplanets*, eds. H. J. Deeg, & J. A. Belmonte (Berlin: Springer), 143
- Mordasini, C., Alibert, Y., & Benz, W. 2009, *A&A*, **501**, 1139
- Mordasini, C., Alibert, Y., Georgy, C., et al. 2012a, *A&A*, **547**, A112
- Mordasini, C., Alibert, Y., Klahr, H., & Henning, T. 2012b, *A&A*, **547**, A111
- Mulders, G. D., Pascucci, I., Apai, D., & Ciesla, F. J. 2018, *AJ*, **156**, 24
- Mulders, G. D., O'Brien, D. P., Ciesla, F. J., Apai, D., & Pascucci, I. 2020 *ApJ*, **897**, 72
- Murchikova, L., & Tremaine, S. 2020, *AJ*, **160**, 160
- Netto, Y., Lorenzo-Oliveira, D., Meléndez, J., et al. 2021, *AJ*, **162**, 160
- Oliphant, T. E. 2006, *A Guide to NumPy* (USA: Trelgol Publishing), 1
- Paardekooper, S. J., Baruteau, C., & Kley, W. 2011, *MNRAS*, **410**, 293
- pandas development team. T. 2020, pandas-dev/pandas: Pandas
- Pepe, F., Correia, A. C. M., Mayor, M., et al. 2007, *A&A*, **462**, 769
- Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, *Icarus*, **124**, 62
- Quanz, S. P., Ottiger, M., Fontanet, E., et al. 2022, *A&A*, **664**, A21
- Rauer, H., Catala, C., Aerts, C., et al. 2014, *Exp. Astron.*, **38**, 249
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, *J. Astron. Teles. Instrum. Syst.*, **1**, 014003
- Rodríguez, J. E., Becker, J. C., Eastman, J. D., et al. 2018, *AJ*, **156**, 245
- Rojas-Ayala, B., Covey, K. R., Muirhead, P. S., & Lloyd, J. P. 2012, *ApJ*, **748**, 93
- Rosenthal, L. J., Fulton, B. J., Hirsch, L. A., et al. 2021, *ApJS*, **255**, 8
- Safsten, E. D., Dawson, R. I., & Wolfgang, A. 2020, *AJ*, **160**, 214
- Sahlmann, J., Lazorenko, P. F., Ségransan, D., et al. 2016, *A&A*, **595**, A77
- Santos, N. C., Bouchy, F., Mayor, M., et al. 2004, *A&A*, **426**, L19
- Sarkar, S. 2022, *MNRAS*, **512**, 5228
- Schib, O., Mordasini, C., Wenger, N., Marleau, G. D., & Helled, R. 2021, *A&A*, **645**, A43
- Schlecker, M., Mordasini, C., Emsenhuber, A., et al. 2021a, *A&A*, **656**, A71
- Schlecker, M., Pham, D., Burn, R., et al. 2021b, *A&A*, **656**, A73
- Scott, D. W. 2015, *Multivariate Density Estimation: Theory, Practice, and Visualization* (Hoboken: Wiley)
- Seager, S., Kuchner, M., Hier-Majumder, C. A., & Militzer, B. 2007, *ApJ*, **669**, 1279
- Seager, S., Knapp, M., Demory, B.-O., et al. 2021, *AJ*, **161**, 117
- Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, **24**, 337
- Shallue, C. J., & Vanderburg, A. 2018, *AJ*, **155**, 94
- Sharma, R., Gupta, M., & Kapoor, G. 2010, *J. Math. Inequalities*, **4**, 355
- Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S. 2010, *Nature*, **465**, 1049
- Staab, D., Haswell, C. A., Barnes, J. R., et al. 2020, *Nat. Astron.*, **4**, 399
- Standish, E. M. 1992, SSD JPL NASA, 1
- Stassun, K. G., Collins, K. A., & Gaudi, B. S. 2017, *AJ*, **153**, 136
- Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, *AJ*, **158**, 138
- Sun, L., Ioannidis, P., Gu, S., et al. 2019, *A&A*, **624**, A15
- Swastik, C., Banyal, R. K., Narang, M., et al. 2021, *AJ*, **161**, 114
- Thiabaud, A., Marboeuf, U., Alibert, Y., et al. 2014, *A&A*, **562**, A27
- Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, *ApJS*, **235**, 38
- Trifonov, T., Kürster, M., Zechmeister, M., et al. 2018, *A&A*, **609**, A117
- Tychoniec, Ł., Tobin, J. J., Karska, A., et al. 2018, *ApJS*, **238**, 19
- Udry, S., Bonfils, X., Delfosse, X., et al. 2007, *A&A*, **469**, L43
- Udry, S., Dumusque, X., Lovis, C., et al. 2019, *A&A*, **622**, A37
- Van Rossum, G., & Drake, F. L. 2009, *Python 3 Reference Manual* (Scotts Valley, CA: CreateSpace)
- Vanderburg, A., Becker, J. C., Buchhave, L. A., et al. 2017, *AJ*, **154**, 237
- Veras, D., & Armitage, P. J. 2004, *MNRAS*, **347**, 613
- Vidal-Madjar, A., Lecavelier des Etangs, A., Désert, J. M., et al. 2003, *Nature*, **422**, 143
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Meth.*, **17**, 261
- Vogt, S. S., Burt, J., Meschiari, S., et al. 2015, *ApJ*, **814**, 12
- Vogt, S. S., Butler, R. P., Burt, J., et al. 2017, *AJ*, **154**, 181
- Wang, H. S., Lineweaver, C. H., & Ireland, T. R. 2018, *Icarus*, **299**, 460
- Wang, S. 2017, *Res. Notes Am. Astron. Soc.*, **1**, 26
- Waskom, M., & the seaborn development team. 2020, <https://github.com/mwaskom/seaborn>
- Weiss, L. M., & Petigura, E. A. 2020, *ApJ*, **893**, L1
- Weiss, L. M., Marcy, G. W., Rowe, J. F., et al. 2013, *ApJ*, **768**, 14
- Weiss, L. M., Marcy, G. W., Petigura, E. A., et al. 2018, *AJ*, **155**, 48
- Weiss, L. M., Dai, F., Huber, D., et al. 2021, *AJ*, **161**, 56
- Yoffe, G., Ofir, A., & Aharonson, O. 2021, *ApJ*, **908**, 114
- Zhu, W. 2020, *AJ*, **159**, 188
- Zhu, W., & Dong, S. 2021, *ARA&A*, **59**, 291

Appendix A: Bern Model: Additional details

In this section, we provide some additional details on the physics included in the Bern model and how it is utilised to simulate synthetic planetary systems. Finally, we give an overview on comparisons between the output of the Bern Model and observed planetary systems. For the historic development, we refer to [Alibert et al. \(2004, 2005\)](#); [Mordasini et al. \(2009\)](#); [Alibert et al. \(2011\)](#); [Mordasini et al. \(2012a,b\)](#); [Alibert et al. \(2013\)](#); [Fortier et al. \(2013\)](#); [Marboeuf et al. \(2014b\)](#); [Thiabaud et al. \(2014\)](#); [Dittkrist et al. \(2014\)](#); [Jin et al. \(2014\)](#) and reviews in [Benz et al. \(2014\)](#); [Mordasini \(2018\)](#).

The Bern model is based on the core accretion paradigm of planetary formation ([Pollack et al. 1996](#)). The model includes stellar evolution for a solar-mass star, using evolution tracks from [Baraffe et al. \(2015\)](#). The star interacts with the protoplanetary disk and influences its thermodynamical properties. The protoplanetary disk has two phases: gas and solid. We model this disk using the approaches of viscous angular momentum transport ([Lynden-Bell & Pringle 1974](#); [Veras & Armitage 2004](#); [Hueso & Guillot 2005](#)). Turbulence is characterised by the [Shakura & Sunyaev \(1973\)](#) approach, with $\alpha = 2 \times 10^{-3}$. Gas from the disk is accreted by planets, host star, and lost via photo-evaporation. The 1D geometrically thin disk evolution is studied up to 1000 au. The initial mass of this gas disk and its lifetime are randomly drawn for each run of the simulation. The solid phase of the disk is composed of a swarm of planetesimals. The solid disk is modelled as a fluid which evolves via (a) accretion by growing planets; (b) interaction with the gaseous disk; (c) dynamical stirring from planets and other planetesimals; and so on ([Fortier et al. 2013](#)). The initial mass of the solid disk depends on the metallicity of the star and also on the condensation state of the molecules in the disk ([Thiabaud et al. 2014](#)). The host star metallicity is randomly drawn for each run of the simulation.

We added 100 protoplanetary embryos to the protoplanetary disk. The initial location of each embryo was varied from one simulation to another. It was also ensured that no two embryos start within 10 Hill radii of each other ([Kokubo & Ida 1998, 2002](#)). Embryos accrete from their feeding zones and any overlap may lead to competition ([Alibert et al. 2013](#)). The accretion rate depends on the collision probability between a protoplanet and a planetesimal, which in turn is influenced by the dynamical state of the solid disk.

Gas accretion occurs in several phases ([Mordasini et al. 2012b](#)). Initially, the gas disk smoothly transitions as a gaseous envelope around all planets – the attached phase. For planets that are massive enough to undergo runaway gas accretion, the rate of gas supply from the disk may not be enough. In these scenarios, the planet detaches from the gas disk and rapidly contracts to R_J . After the gas disk dissipates, all planets are in the isolated phase. Gas accretion from the disk is no longer possible and in this phase, the planets contract and cool. For all the planets, their internal structure is modelled at each time step. We assume planets are spherically symmetric and composed of accreted materials that arranges itself in layers: iron core, silicate mantle, water ice, and H/He gaseous envelope (if accreted).

Next, we use these recipes to simulate several thousands of planetary systems in an approach called population synthesis ([Emsenhuber et al. 2021b](#)). We start 1000 star-disk-embryo systems with some fixed as well as some randomly drawn properties. The initial properties are inspired by observations of disks [Tychoniec et al. \(2018\)](#). We, then, numerically modelled these systems, endowing them with additional physics at the same time. Numerically, we incorporated multi-body dynamical interactions via N -body simulations. Planet-disk interactions leading

to orbital migration and eccentricity and inclination damping were also incorporated in the N -body [Coleman & Nelson \(2014\)](#); [Paardekooper et al. \(2011\)](#); [Dittkrist et al. \(2014\)](#). We followed these numerically intensive steps for 20 Myrs and then stopped the N -body calculations. The model then continued to evaluate the internal structure of all planets in the system for 10 Gyrs.

The recent version of these simulations has been published in the New Generation Planetary Population Synthesis (NGPPS) series of papers ([Emsenhuber et al. 2021a,b](#); [Schlecker et al. 2021a,b](#); [Burn et al. 2021](#); [Mishra et al. 2021](#)). The output of these models have been compared with observations in several works. [Draskowska et al. \(2022\)](#) compares the occurrence rates of synthetic systems with observations. [Schlecker et al. \(2021a\)](#) studies the warm Super Earth and cold Jupiter correlation in the synthetic systems. [Mishra et al. \(2021\)](#) analyse the 'peas in a pod' architecture and compare synthetic systems with observations from [Weiss et al. \(2018\)](#). [Mulders et al. \(2018\)](#) present a detailed comparison of the synthetic models with Kepler observations.

Appendix B: Stellar and planetary data references

1. Sun: [Archinal et al. \(2018\)](#); [Standish \(1992\)](#); [Wang et al. \(2018\)](#); [Helffrich \(2017\)](#); [Jacobson et al. \(2006\)](#); [Jacobson \(2014, 2009\)](#)
2. Trappist-1: [Agol et al. \(2021\)](#); [Gillon et al. \(2017\)](#); [Burgasser & Mamajek \(2017\)](#); [Grimm et al. \(2018\)](#)
3. TOI-178: [Leleu et al. \(2021\)](#)
4. HD 10180: [Lovis et al. \(2011\)](#); [Kane & Gelino \(2014\)](#)
5. HD 219134: [Seager et al. \(2021\)](#); [Bonfanti & Gillon \(2020\)](#); [Vogt et al. \(2015\)](#)
6. HD 34445: [Vogt et al. \(2017\)](#)
7. Kepler-11: [Berger et al. \(2020\)](#); [Lissauer et al. \(2013\)](#)
8. Kepler-20: [Fressin et al. \(2011\)](#); [Buchhave et al. \(2016\)](#)
9. Kepler-80: [MacDonald et al. \(2016\)](#); [Shallue & Vanderburg \(2018\)](#)
10. K2-138: [Lopez et al. \(2019\)](#)
11. 55 Cnc: [Bourrier et al. \(2018\)](#)
12. GJ 667 C: [Anglada-Escudé et al. \(2013\)](#)
13. HD 158259: [Hara et al. \(2020\)](#); [Gáspár et al. \(2016\)](#)
14. HD 40307: [Díaz et al. \(2016\)](#); [Stassun et al. \(2019\)](#)
15. Kepler-102: [Berger et al. \(2020\)](#); [Marcy et al. \(2014\)](#)
16. Kepler-33: [Berger et al. \(2020\)](#); [Lissauer et al. \(2012\)](#); [Hadden & Lithwick \(2017\)](#)
17. Kepler-62: [Berger et al. \(2020\)](#); [Borucki et al. \(2013\)](#)
18. HD 20781: [Udry et al. \(2019\)](#)
19. TOI-561: [Lacedelli et al. \(2021\)](#); [Weiss et al. \(2021\)](#)
20. DMPP-1: [Staab et al. \(2020\)](#)
21. GJ 3293: [Astudillo-Defru et al. \(2017\)](#)
22. GJ 676 A: [Sahlmann et al. \(2016\)](#); [Stassun et al. \(2017\)](#)
23. GJ 876: [Trifonov et al. \(2018\)](#); [Rojas-Ayala et al. \(2012\)](#)
24. HD 141399: [Hébrard et al. \(2016\)](#)
25. HD 160691: [Goździewski et al. \(2007\)](#); [Pepe et al. \(2007\)](#)
26. HD 20794: [Goździewski et al. \(2007\)](#); [Pepe et al. \(2007\)](#)
27. HD 215152: [Goździewski et al. \(2007\)](#); [Pepe et al. \(2007\)](#)
28. HR 8799: [Marois et al. \(2008\)](#); [Gravity Collaboration \(2019\)](#); [Swastik et al. \(2021\)](#)
29. K2-266: [Rodríguez et al. \(2018\)](#)
30. K2-285: [Rodríguez et al. \(2018\)](#)
31. Kepler-89: [Berger et al. \(2020\)](#); [Weiss et al. \(2013\)](#)
32. Kepler-106: [Berger et al. \(2020\)](#); [Marcy et al. \(2014\)](#)
33. Kepler-107: [Berger et al. \(2020\)](#); [Bonomo et al. \(2019\)](#)
34. Kepler-223: [Berger et al. \(2020\)](#); [Mills et al. \(2016\)](#)
35. Kepler-411: [Berger et al. \(2020\)](#); [Sun et al. \(2019\)](#)
36. Kepler-48: [Berger et al. \(2020\)](#); [Marcy et al. \(2014\)](#)

37. Kepler-65: [Berger et al. \(2020\)](#); [Mills et al. \(2019\)](#)
38. Kepler-79: [Berger et al. \(2020\)](#); [Yoffe et al. \(2021\)](#)
39. WASP-47: [Vanderburg et al. \(2017\)](#)
40. tau Cet: [Vanderburg et al. \(2017\)](#)
41. HD 164922: [Benatti et al. \(2020\)](#); [Rosenthal et al. \(2021\)](#)

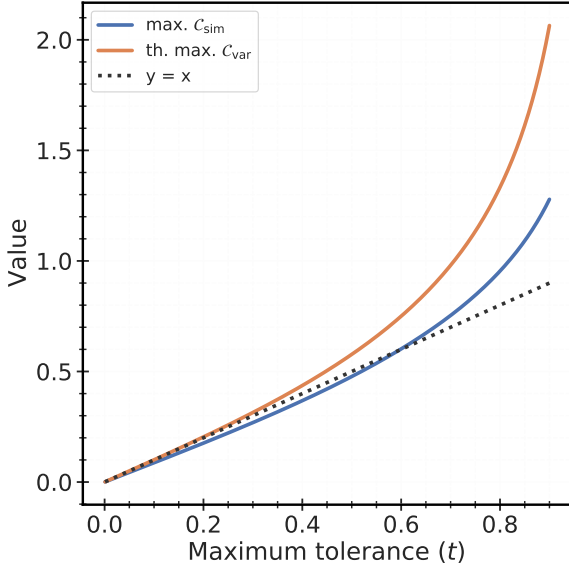


Fig. C.1. Maximum value of the coefficient of similarity (blue) and the theoretical maximum value of the coefficient of variation (orange) is plotted against the maximum tolerance, t .

Appendix C: Derivation of limits

We consider a set Q of quantities q , namely, $Q = \{q_i\}$ where q_i could be the mass, radius or other parameter of a planet, and the index, $i \in [1, n]$, identifies a planet (with 1 being the innermost planet). We assume that all $q_i \geq 0$. The quantities q_i are expressed as:

$$q_i = q' (1 \pm t_i). \quad (\text{C.1})$$

The quantities, q_i , are decomposed around some value q' such that all t_i are minimised; t_i is a measurement of the fractional difference (or tolerance) between q' and q_i . Since all individual tolerances are a positive quantity, they will satisfy the following relation:

$$0 \leq t_i \leq t. \quad (\text{C.2})$$

Appendix C.1. Mean

Let us consider the mean of the quantities, \bar{q}_i :

$$\begin{aligned} \bar{q}_i &= \frac{Q}{n} = \frac{\sum_i q_i}{n} \\ &= \frac{q'}{n} (n \pm t_1 \pm t_2 \pm \dots \pm t_n). \end{aligned} \quad (\text{C.3})$$

The mean takes its maximum value only when all individual t_i values take their maximum and are added up. This gives:

$$\max \bar{q}_i = q' (1 + t). \quad (\text{C.4})$$

Similarly, the minimum value of the mean is:

$$\min \bar{q}_i = q' (1 - t). \quad (\text{C.5})$$

The extreme value of the mean occurs when all the individual quantities are extremised. However, in this scenario, since all quantities are equal, the coefficient of variation is identically 0.

Appendix C.2. Coefficient of similarity

We start with the definition of the coefficient of similarity,

$$C_S(q) = \frac{1}{n-1} \sum_{i=1}^{i=n-1} \left(\log \frac{q_{i+1}}{q_i} \right). \quad (\text{C.6})$$

Inserting Eq. C.1, in the definition, we get:

$$C_S(q) = \frac{1}{n-1} \sum_{i=1}^{i=n-1} \left(\log \frac{1 \pm t_{i+1}}{1 \pm t_i} \right). \quad (\text{C.7})$$

This formulation shows that the coefficient of similarity depends only on the fractional differences (tolerances) between q_i values – and not on their actual values. This is a desirable property.

Next, we evaluate the max C_S as,

$$\begin{aligned} \max C_S(q) &= \max \left[\frac{1}{n-1} \sum_{i=1}^{i=n-1} \left(\log \frac{1 \pm t_{i+1}}{1 \pm t_i} \right) \right], \\ &= \frac{1}{n-1} \max \left[\sum_{i=1}^{i=n-1} \left(\log \frac{1 \pm t_{i+1}}{1 \pm t_i} \right) \right], \\ &= \frac{1}{n-1} \sum_{i=1}^{i=n-1} \log \max \left[\frac{1 \pm t_{i+1}}{1 \pm t_i} \right]. \end{aligned} \quad (\text{C.8})$$

In the first step, we commuted the max operator with the fraction $(n-1)^{-1}$ because we are interested in the maximum for a constant n . Next, knowing that the maximum of a sum occurs at the sum of maximum summands and that \log is a monotonically increasing function, we further commute the max operator.

We observe the following:

$$\max \left[\frac{1 \pm t_{i+1}}{1 \pm t_i} \right] \quad \text{when} \quad \left\{ \begin{array}{l} \pm t_{i+1} \rightarrow +t \\ \pm t_i \rightarrow -t \end{array} \right\}. \quad (\text{C.9})$$

This implies that

$$\begin{aligned} \max C_S(q) &= \log \frac{1+t}{1-t}, \\ \min C_S(q) &= -\max C_S = \log \frac{1-t}{1+t}, \end{aligned} \quad (\text{C.10})$$

where the second equality can be similarly derived. Fig. C.1 shows the variation of max C_S as a function of tolerance t . We note that the limits of the coefficient of similarity do not depend on n , and we verified our results with numerical simulations. ■

Appendix C.3. Coefficient of Variation

We start with the definition of the coefficient of variation,

$$C_V(q) = \frac{\sigma(q)}{\bar{q}}, \quad (\text{C.11})$$

and we note that the minimum value of the coefficient of variation is zero and it occurs when all q_i values are equal, thereby giving no variance.

In the literature, we can find some derivations for the maximum value of the coefficient of variation ([Katsnelson & Kotz 1957](#); [Sharma et al. 2010](#)). [Katsnelson & Kotz \(1957\)](#) show that the upper limit of the coefficient of variation is $\sqrt{n-1}$ when all but one q_i is zero. However, this limit is only a particular case of our formulation (specifically, $q_1 = q'$ and $q_{i \neq 1} = 0$). Here, we derive the limits for a more general scenario.

We consider that:

$$C_V^2 = \frac{1}{n} \sum_{i=1}^{i=n} \underbrace{\left(1 - \frac{q_i}{\bar{q}}\right)^2}_{=A}. \quad (\text{C.12})$$

Here, we have squared the definition of C_V and used the definition of the standard deviation $\sigma(q)$. As an aside, we note that the equation above shows that the coefficient of variation is zero when all $q_i = \bar{q}$, as noted before. We note that the maximum value of C_V^2 occurs when the term A (in parenthesis) is maximised. Denoting $\sum_{i=1} q_i$ by Q , we consider the term in the parenthesis,

$$A = 1 - \frac{nq_i}{Q} = \frac{Q - nq_i}{Q}. \quad (\text{C.13})$$

The condition for the general maxima of the coefficient of variation, in our formulation, is when one of the quantity (say q_1 takes the largest possible value, while all others take the smallest possible value):

$$\begin{aligned} q_1 &= q'(1+t) \\ q_{i \neq 1} &= q'(1-t). \end{aligned} \quad (\text{C.14})$$

The mean in this scenario becomes (marked with '):

$$\bar{q}'' = \frac{q'(1+t) + (n-1) \times q'(1-t)}{n} = \frac{q'}{n} [n(1-t) + 2t]. \quad (\text{C.15})$$

The variance in this scenario becomes (marked with '):

$$\sigma''^2(q) = \frac{1}{n} \left\{ \underbrace{\left[q'(1+t) - \bar{q}'' \right]^2}_{= 2q't \left(\frac{n-1}{n} \right)} + (n-1) \underbrace{\left[q'(1-t) - \bar{q}'' \right]^2}_{= \frac{-2q't}{n}} \right\}. \quad (\text{C.16})$$

This gives:

$$\sigma''(q) = \left(\frac{2q't}{n} \right) \sqrt{n-1}. \quad (\text{C.17})$$

Finally, the general expression for the maximum value of the coefficient of variation becomes:

$$\max C_V(q) = \frac{\sigma''(q)}{\bar{q}''} = \frac{2t \sqrt{n-1}}{n(1-t) + 2t}. \quad (\text{C.18})$$

This expression recovers the particular case derived in literature when we set $t = 1$. From this expression, we note that the upper limit of the coefficient of variation does not depend on the actual values of the quantities, but it depends on the number of quantities in the set, Q , and the maximum tolerance, t . This new formulation allows us to extract the upper limit of the coefficient of variation for any set whose maximum tolerance, t , is known. Interestingly, the above expression gives appropriate result when absurd inputs are considered. For example, when there are no planets in a system, $\max C_V|_{n=0} = \sqrt{-1}$, and when there is only one planet in a system, $\max C_V|_{n=1} = 0$. For a system of two planets, the upper limit is exactly the fractional difference (or tolerance), that is, $\max C_V|_{n=2} = t$.

Furthermore, varying over n , and assuming $t \in [0, 1)$, allows us to derive the theoretical maximum possible value for the coefficient of variation. This occurs at $n = \frac{2}{1-t}$ and gives:

$$\max C_V(q) \Big|_{n=\frac{2}{1-t}}(q) = \frac{t}{\sqrt{1-t^2}}. \quad (\text{C.19})$$

Figure C.1 shows the variation of the theoretical max, C_V , as a function of tolerance t . ■

Appendix D: Classification boundaries for architectures classes

In this section, we present some considerations that motivate the boundaries between the four architecture classes for planetary masses. In the current formulation (Eq. 3), there are two boundaries that need to be identified. We deal with the distinction between similar and mixed class first, and then distinguish ordered/anti-ordered architecture classes.

Appendix D.1. Similar versus mixed

We saw in Sect. 3.2, it is difficult to distinguish between mixed and similar architecture classes using the coefficient of similarity alone. Mixed systems are characterised by large increasing or decreasing variations, which may cancel each other out and lead to small values of $C_S(M)$. Nevertheless, the coefficient of variation can distinguish between large variations in values. Figure D.1 shows the $C_V(M)$ as a function of the number of planets in a planetary system. The left panel shows all synthetic systems from the Bern model, while we only show systems with $|C_S(M)| \leq 0.2$ in the right panel.

Clearly, there are two clusters of planetary systems. The cluster on the lower right-hand side corresponds to similar class systems. Mixed systems, having large values of $C_V(M)$, are spread over the top left region. It is clear that the boundary between similar and mixed classes depends on the number of planets. The black line (corresponding to $y = \frac{\sqrt{n-1}}{2}$) neatly separates the two clusters. We have chosen this equation to disentangle similar architectures from the mixed class. This equation has, incidentally, two key properties: 1) it ensures that no two planet system can be of mixed architecture and 2) it happens to be exactly half of the maximum possible value of the coefficient of variation.

Appendix D.2. Ordered and anti-ordered

Having motivated the boundary between similar and mixed class, we are now left with three groupings of architecture classes. These three groupings correspond to $C_S(M) \ll 0$ (anti-ordered), $C_S(M) \sim 0$ (similar/mixed), and $C_S(M) \gg 0$ (ordered). This suggests that we require two boundaries to distinguish these three groups. However, we posit that the boundaries between ordered and anti-ordered should be symmetric around 0. Thus, we are left with only one boundary.

ordered (or anti-ordered) systems differ in their architecture from similar/mixed classes in that the quantity (mass here) continues to show an increasing (or decreasing) trend with distance. For all planetary systems in the Bern model, we measure the Spearman correlation coefficient, R , between the planetary masses and their distance from the host star. The Spearman R , measuring the monotonicity between two datasets, varies from -1 to +1, with 0 indicating no correlation. A positive correlation implies that as x increases, so does y . Negative correlations imply that as x increases, y decreases.

Figure D.1 shows the $C_S(M)$ of synthetic systems as function of their Spearman correlation R (mass and distance). We note that there is a large cluster of points towards $C_S(M) \sim 0$. This group corresponds to the similar and mixed architecture classes. There are some points to the top right (including those with $R = +1$ – corresponding to planetary systems in which planetary masses are monotonically increasing with distance). There is a scatter of points towards the bottom-left (including some systems with $R = -1$).

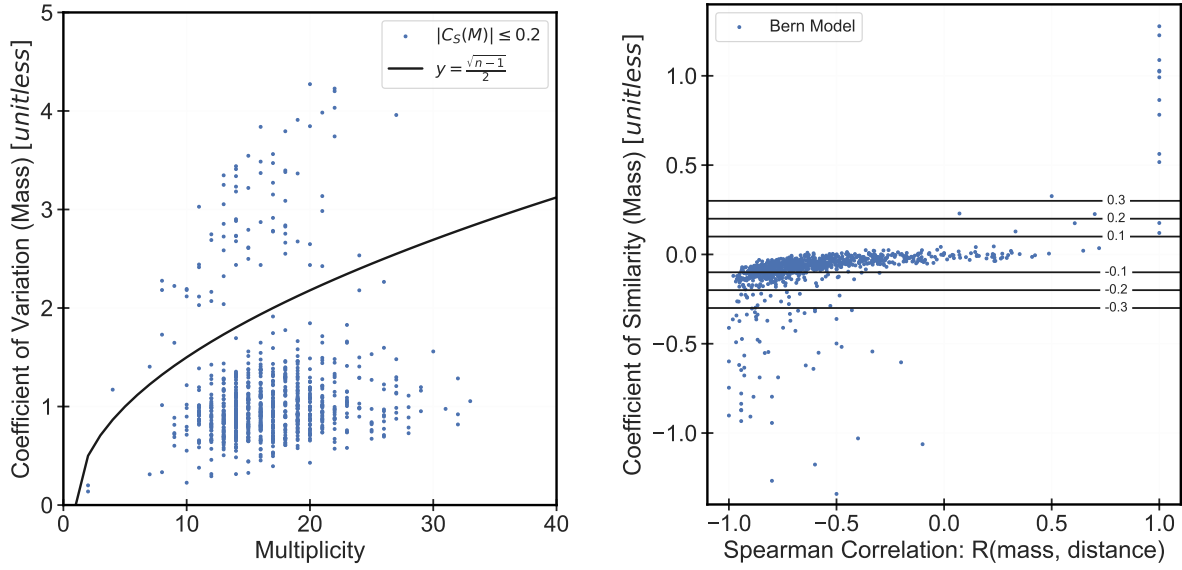


Fig. D.1. Classification boundaries for architecture classes. *Left:* Boundary between similar and mixed class. The panel show the coefficient of variation for synthetic planetary systems as a function of the number of planets in a system for systems with $|C_S(M)| \leq 0.2$. Two clusters are clearly distinguishable, allowing us to fix the boundary between the similar and mixed architecture classes. *Right:* Boundary between ordered and anti-ordered. This plot shows the coefficient of similarity of synthetic planetary systems as a function of the Spearman correlation coefficient between the planetary masses and distances of that system. Thick horizontal lines correspond to potential boundaries.

First, we note that the comparison of the coefficient of similarity with Spearman R fulfils some expectation. For example, there are no points in bottom-right or top-left sections of this plot. Second, our objective is to isolate the central cluster of points from all other scattered points. We note that $|C_S(M)| = 0.1$ fails as a boundary, since it does not include the full central cluster. Both $|C_S(M)| = 0.2$ and 0.3 could succeed. Going beyond, a value of 0.3 would add many unnecessary points to the central cluster.

To further motivate our choice of boundary, namely, $|C_S(M)| = 0.2$, we show the mass-distance diagram of 12 randomly selected systems with $-0.3 < C_S(M) < -0.2$ (out of 19) in Fig. D.2. We note that all systems show the qualitative features of an anti-ordered system, namely, massive planets in the inner region and small planets in the outer region. Since all of these planets have their $C_S(M) < -0.2$, we use $|C_S(M)| = 0.2$ as a boundary between ordered, anti-ordered, and similar+mixed architecture classes. Future works may explore improvements to our selected boundaries using additional ideas from K-means or hierarchical clusterings.

Appendix E: A gallery of architecture types: Mass-distance diagrams

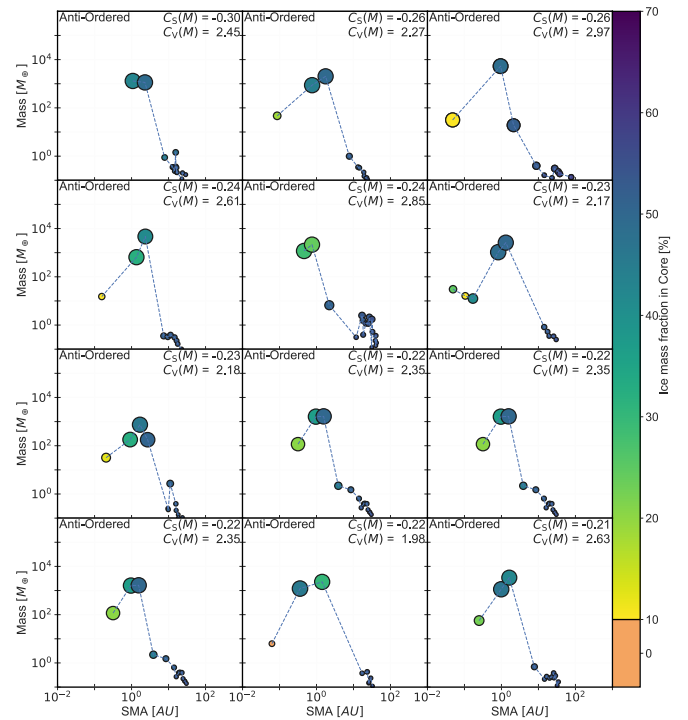


Fig. D.2. Mass-distance diagram. This plot shows the planetary masses as a function of distance for some planetary systems with $-0.3 < C_S(M) < -0.2$. The dashed line connects that planets in the system and serves to highlight the arrangement and distribution of masses. The size of each circle corresponds to the planet's radius and the colour of each planet also shows its core water mass fraction.

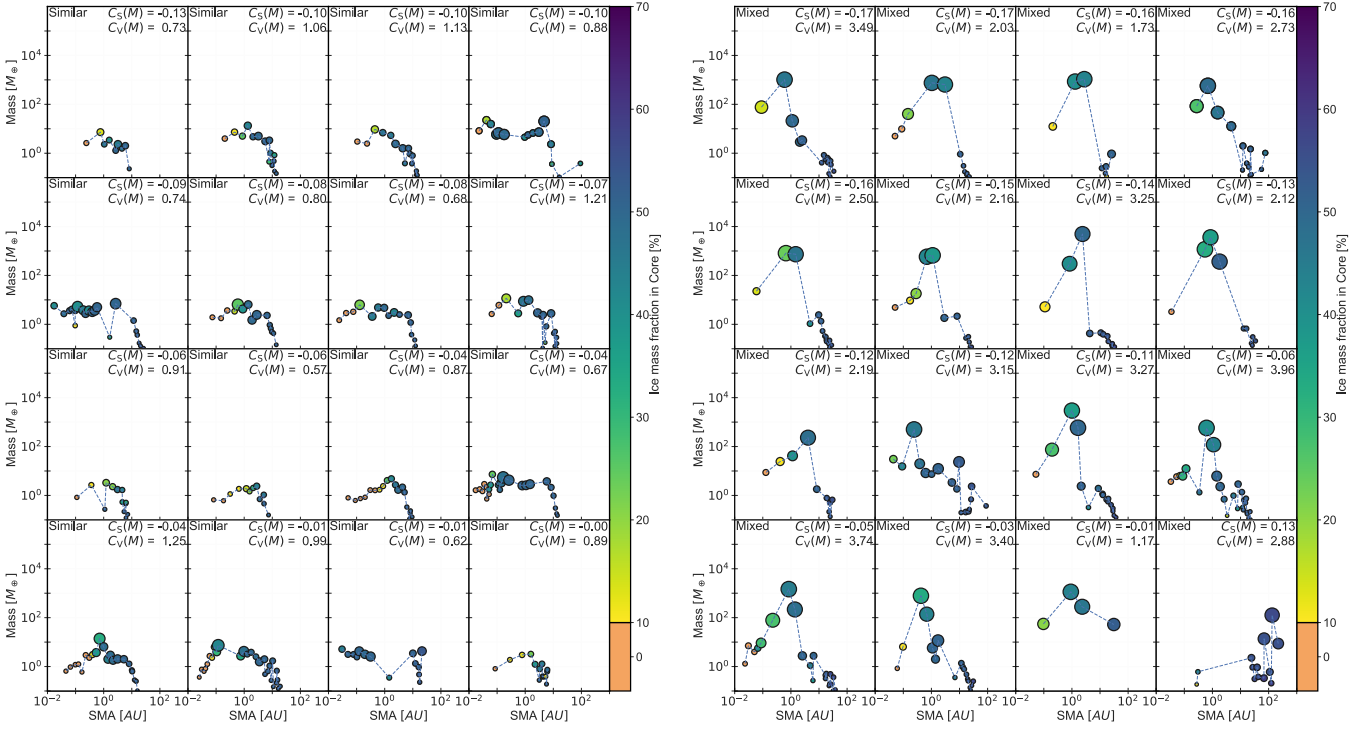


Fig. E.1. A gallery of planetary system architectures. These plots show the mass-distance diagram for similar (left) and mixed (right) planetary systems from the Bern Model. Each circle represents a planet, its size corresponds to the planetary radius, and its colour represents the fraction of ice in the planetary core. Each panel shows the $C_S(M)$ as well as the $C_V(M)$ of the system.

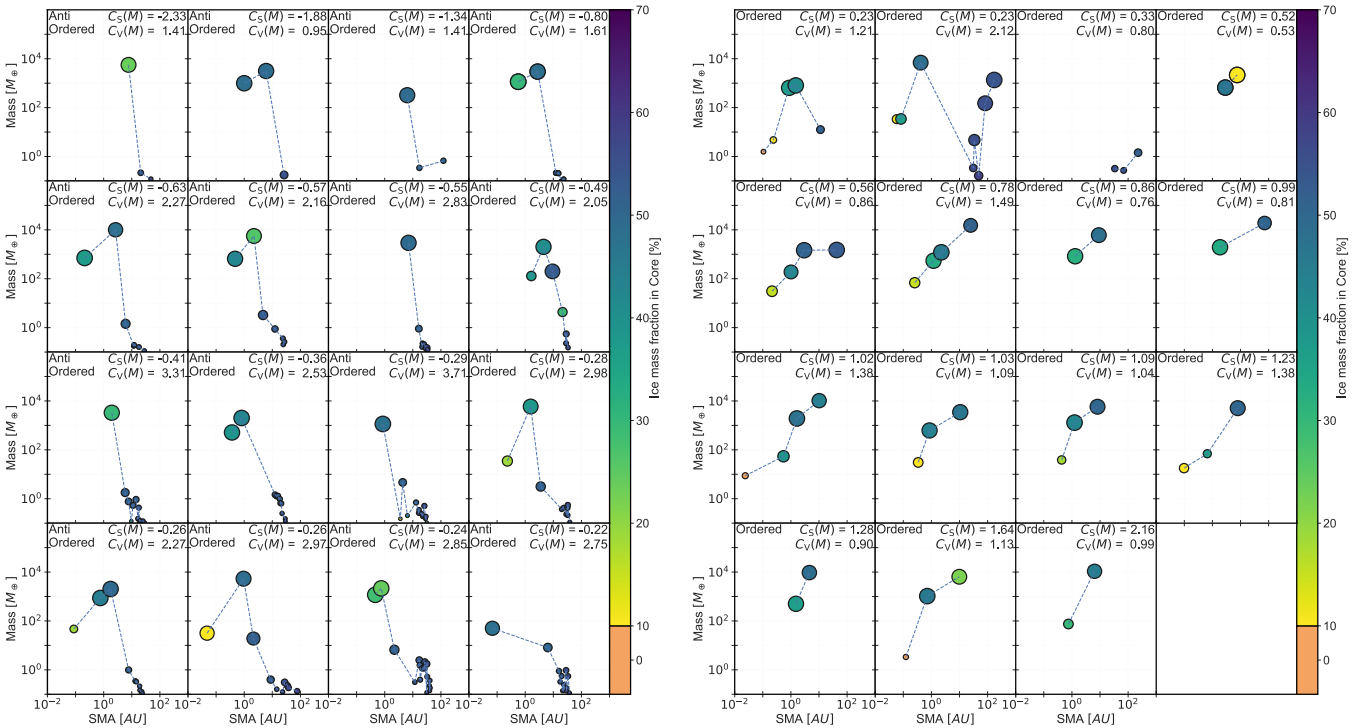


Fig. E.2. A gallery of planetary system architectures. These plots show the mass-distance diagram for anti-ordered (left) and ordered (right) planetary systems from the Bern Model. Each circle represents a planet, its size corresponds to the planetary radius, and its colour represents the fraction of ice in the planetary core. Each panel shows the $C_S(M)$ as well as the $C_V(M)$ of the system.