Predicting the long-term stability of compact multiplanet systems

Daniel Tamayo[a,1], Miles Cranmer[a], Samuel Hadden[b], Hanno Rein[c,d], Peter Battaglia[e], Alysa Obertas[d,f], Philip J. Armitage[g,h], Shirley Ho[h,a,i], David N. Spergel[h], Christian Gilbertson[j], Naireen Hussain[d], Ari Silburt[j,c,d], Daniel Jontof-Hutter[k], and Kristen Menou[c,d,l]

[a]Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544; [b]Center for Astrophysics, Harvard & Smithsonian, Cambridge, MA 02138; [c]Department of Physical and Environmental Sciences, University of Toronto at Scarborough, Toronto, ON M1C 1A4, Canada; [d]David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada; [e]Google DeepMind, London N1C 4AG, United Kingdom; [f]Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, ON M5S 3H8, Canada; [g]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11790; [h]Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010; [i]Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15217; [j]Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802; [k]Department of Physics, University of the Pacific, Stockton, CA 95211; and [l]Department of Physics, University of Toronto, Toronto, ON M5S 1A7, Canada

Edited by Christopher F. McKee, University of California, Berkeley, CA, and approved June 15, 2020 (received for review January 22, 2020)

We combine analytical understanding of resonant dynamics in two-planet systems with machine-learning techniques to train a model capable of robustly classifying stability in compact multiplanet systems over long timescales of 10^9 orbits. Our Stability of Planetary Orbital Configurations Klassifier (SPOCK) predicts stability using physically motivated summary statistics measured in integrations of the first 10^4 orbits, thus achieving speed-ups of up to 10^5 over full simulations. This computationally opens up the stability-constrained characterization of multiplanet systems. Our model, trained on ∼100,000 three-planet systems sampled at discrete resonances, generalizes both to a sample spanning a continuous period-ratio range, as well as to a large five-planet sample with qualitatively different configurations to our training dataset. Our approach significantly outperforms previous methods based on systems' angular momentum deficit, chaos indicators, and parametrized fits to numerical integrations. We use SPOCK to constrain the free eccentricities between the inner and outer pairs of planets in the Kepler-431 system of three approximately Earth-sized planets to both be below 0.05. Our stability analysis provides significantly stronger eccentricity constraints than currently achievable through either radial velocity or transit-duration measurements for small planets and within a factor of a few of systems that exhibit transit-timing variations (TTVs). Given that current exoplanet-detection strategies now rarely allow for strong TTV constraints [S. Hadden, T. Barclay, M. J. Payne, M. J. Holman, *Astrophys. J.* 158, 146 (2019)], SPOCK enables a powerful complementary method for precisely characterizing compact multiplanet systems. We publicly release SPOCK for community use.

exoplanets | chaos | machine learning | orbital dynamics | dynamical systems

Isaac Newton, having formulated his law of gravitation, recognized that it left the long-term stability of the Solar System in doubt. Would the small near-periodic perturbations the planets exert on one another average out over long timescales, or would they accumulate until orbits cross, rendering the system unstable to planetary collisions or ejections?

The central difficulty arises from the existence of resonances, where there is an integer ratio commensurability between different frequencies in the system. These resonances complicate efforts to average the dynamics over long timescales. Work on this problem culminated in the celebrated Komolgorov–Arnold–Moser (KAM) theorem (1–3), which guarantees the existence of stable, quasiperiodic trajectories below a specified perturbation strength. Unfortunately, the KAM theorem is generally not informative in this context, since it typically can only guarantee stability for masses far below the planetary regime (4, 5).

Without a clear path to a full solution, we focus on the limit of closely separated planets. This regime has important applications for understanding the orbital architectures of planetary systems beyond our own (exoplanets), since strong observational biases toward detecting planets close to their host star result in compact populations of observed multiplanet systems.* In these dynamically delicate configurations, it is possible for most of the orbital solutions inferred from noisy data to undergo violent dynamical instabilities when numerically integrated forward in time for even 0.1% of the system's age (6, 7). Since one does not expect to discover most systems just prior to such a cataclysm, this offers an opportunity to constrain the masses and orbital parameters of such planets by rejecting configurations that lead to rapid instability. In this way, previous authors have performed direct numerical (N-body) integrations to narrow down physical orbital architectures and formation histories for important exoplanet discoveries (e.g., refs. 8–12).

Significance

Observations of planets beyond our solar system (exoplanets) yield uncertain orbital parameters. Particularly in compact multiplanet systems, a significant fraction of observationally inferred orbital configurations can lead to planetary collisions on timescales that are short compared with the age of the system. Rejection of these unphysical solutions can thus sharpen our view of exoplanetary orbital architectures. Long-term stability determination is currently performed through direct orbital integrations. However, this approach is computationally prohibitive for application to the full exoplanet sample. By speeding up this process by up to five orders of magnitude, we enable precise exoplanet characterization of compact multiplanet systems and our ability to examine the stability properties of the multiplanet exoplanet sample as a whole.

Author contributions: D.T., M.C., S. Hadden, H.R., P.B., and K.M. designed research; D.T., M.C., A.O., C.G., N.H., and A.S. performed research; S. Hadden, H.R., and P.B. contributed new reagents/analytic tools; D.T., M.C., A.O., C.G., N.H., A.S., and D.J.-H. analyzed data; and D.T., M.C., S. Hadden, H.R., P.J.A., S. Ho, D.N.S., and K.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Zenodo repository (https://zenodo.org/record/3723292). Open-source packages, documentation, and examples are available in GitHub at https://github.com/dtamayo/spock.

[1]To whom correspondence may be addressed. Email: dtamayo@astro.princeton.edu.

First published July 16, 2020.

*We henceforth define compact 3+ planet systems as having at least one trio of planets with period ratios between adjacent planets <2. This currently represents ∼40% of all observed 3+ planet systems.

However, given the high dimensionality of parameters, the computational expense of such long-term integrations typically results in only a small fraction of candidate orbital configurations being explored and integration timespans being many orders of magnitude shorter than the typical Gyr ages of such systems (e.g., refs. 6 and 11–17). This renders the widespread application of such constraints to the ever-growing exoplanet sample orders of magnitude beyond computational reach.

Extensive previous work has narrowed down the particular resonances responsible for dynamical instabilities in compact systems. In particular, analytical studies of tightly spaced two-planet systems (18–20) have shown that the chaos is driven specifically by the interactions between mean motion resonances (MMRs), i.e., integer commensurabilities between planets' orbital periods. The limited number of available MMRs in two-planet systems implies that for initially circular orbits, there exists a critical, mass-dependent separation between the two bodies. For planetary separations below this limit, MMRs are close enough to one another in phase space to overlap and drive rapid instabilities (18, 19), and there is a sharp transition to long-lived configurations beyond it. This result has recently been generalized for eccentric orbits (20).

By contrast in 3+ planet systems, instabilities can occur for separations between adjacent planet pairs significantly beyond the above two-planet limit, and instability times exhibit a continuous and much larger dynamic range (21). Previous work has argued that this cannot be solely explained by the larger number of available MMRs between all possible pairs of planets (22, 23). These authors argue that three-body resonances, i.e., integer combinations between the periods of three bodies are responsible for "filling in the space" between two-body MMRs and driving instabilities over a continuous range of separations.

However, while a clearer physical picture is emerging, theoretical estimates cannot yet quantitatively match the results from numerical integrations (22). Many previous numerical studies have instead presented empirical fits to the overall steep rise in instability times with interplanetary separation, recorded from large suites of numerical integrations (21, 24–30). This is a useful approach for elucidating the underlying dynamics and scalings with dominant parameters but typically involves simplifications such as equal-mass, or equal-separation planets. This limitation, together with modulations near MMRs on overall trends in instability times of up to five orders of magnitude (31), leads to quantitative disagreements between such studies and renders them inadequate for accurately characterizing real multiplanet systems (*Results*).

Here, we present a machine-learning model that can reliably classify the stability of compact 3+ planet configurations over 10^9 orbits. Our model, the Stability of Planetary Orbital Configurations Klassifier (SPOCK), is up to 10^5 times faster than direct integration, computationally opening up the stability constrained characterization of compact multiplanet systems.

Previous Models

Previous numerical efforts to predict the instability times of various orbital configurations can roughly be broken down into four groups.

N-Body. The most straightforward (and computationally costly) method is to run a direct numerical integration. A 10^9 orbit integration with a timestep of 3.5% of the innermost planet's orbital period takes $\sim$7 central processing unit (CPU) hours on a 2.1-GHz Intel Xeon Silver 4116 using the WHFast integrator (32).

Interestingly, even this answer will not be perfect. The fact that planetary systems are chaotic means that a given initial condition should not be considered to have a single instability time. Rather, an N-body integration can be interpreted as sampling a single instability time from a broader distribution of values. If one numerically characterizes the distribution of these instability times, one finds that, for compact systems destabilizing within 10^9 orbits, they are approximately log-normally distributed, with a uniform SD of $\sim$0.4 decades (33, 34). To empirically quantify this fundamental limit to predictability, for each of the integrations in our training dataset, we have run a second "shadow integration" of the same initial conditions offset by one part in 10^{11}. This represents an independent draw from that initial condition's instability time distribution. There will thus be cases where one integration says the configuration is stable, while the other one does not. The existence of these uncertain outcomes sets the fundamental limit any stability classifier can hope to reach.

Hill. Several previous studies have fit functional forms to instability times recorded in large suites of N-body integrations (e.g., refs. 21, 25, 27, 28, and 31). They found that instability times rise steeply with increasing interplanetary separation measured in mutual Hill radii, i.e., the characteristic radius around the planets in which their gravity dominates that of the star (see also refs. 22 and 35),

$$R_H = a_i \left(\frac{m_i + m_{i+1}}{M_\star} \right)^{1/3}, \quad [1]$$

where a_i is the semimajor axis of the inner planet in the pair, m_i and m_{i+1} are the respective planet masses, and $M_\star$ is the stellar mass.[†] While this provides insight into the underlying dynamics (22, 35), other orbital parameters also strongly influence stability. Follow-up studies have considered the effects of finite eccentricities and inclinations (e.g., refs. 24, 26, 36, and 37) but make various simplifying assumptions (e.g., equal interplanetary separations and eccentricities). Different assumptions lead to quantitative disagreements between different studies, and the reliability of their predictions to real systems, where all planets have independent orbital parameters, is unclear.

Angular Momentum Deficit. A classical result in orbital dynamics is that if the effects of MMRs are removed, then planets will exchange angular momenta at fixed semimajor axes (38). Instabilities can still arise under these so-called secular dynamics, through chaos introduced by the overlap of resonances between the slower set of frequencies at which the orbits and their corresponding orbital planes precess (39, 40). In this approximation, there is a conserved quantity (41, 42), termed the Angular Momentum Deficit (AMD). The AMD acts as a constant reservoir of eccentricity and inclination that the planets can exchange among one another. If the AMD is too small to allow for orbit crossing and collisions even in the worst case where all of the eccentricity is given to one adjacent pair of planets, the system is AMD stable (43, 44). This is a powerful and simple analytic criterion, but it has two important caveats. First, because it is a worst-case-scenario estimate, it yields no information on instability timescales for AMD unstable systems. For example, the Solar System is AMD unstable, but most integrations ($\sim$99%) of the Solar System nevertheless remain stable over the Sun's main sequence lifetime (45). Second, the assumed secular model of the dynamics ignores the effects of MMRs, which for closely packed systems are typically nearby (e.g., ref. 46), and are an important source of dynamical chaos (for a generalization of AMD stability in the presence of MMRs in the two-planet case, see ref. 47).

[†]We note that the Hill-sphere scales as the planet–star mass ratio μ to the one-third power. Other authors (e.g., refs. 20, 22, and 35) argue that a $\mu^{1/4}$ scaling is better motivated. These scalings are close to one another, and given the poor performance of such models (Results), we do not pursue this possible correction.

Mean Exponential Growth Factor of Nearby Orbits. Several authors have also used chaos indicators numerically measured from short integrations as a proxy for instability (46, 48). This is appealing given that systems that go unstable typically exhibit chaotic dynamics on shorter timescales. A widely used chaos indicator is the Mean Exponential Growth Factor of Nearby Orbits (MEGNO) (49). However, a planetary system can be chaotic yet never develop destructive instabilities on astrophysically relevant timescales. Additionally, and most importantly, short integrations will fail to measure chaos on timescales longer than those simulated, potentially misclassifying systems that destabilize on long timescales.

Predicting Long-Term Stability

Point-source Newtonian gravity is scale-invariant. We exploit this fact by expressing all masses relative to that of the central star and all times and distances in units of the innermost planet's orbital period and semimajor axis, respectively.

Nondimensionalizing timescales in this way is important when comparing systems with different absolute ages. For example, the $\sim$40 My age of the HR 8799 planets, with an innermost orbital period of $\sim$40 y, only represents 10^6 orbits (10). For these short timescales, numerical integrations[‡] are within reach, and SPOCK is not needed (10).

However, young multiplanet systems with long orbital periods are currently exceedingly rare in the exoplanet sample. Population statistics and strong observational biases result in a multiplanet sample predominantly with innermost orbital periods of $\sim$0.01 to 0.1 y, around stars that are several billion years old. We are thus most often interested in stability over timescales of 10^{11} to 10^{12} orbits, which are computationally prohibitive for the number of candidate orbital configurations that typically require evaluation.

One approach would be to frame the task as a regression problem and predict an instability time for a given initial configuration. However, given that most systems have large dynamical ages $> 10^{11}$ orbits, for many applications, one is simply interested in a binary classification between short-lived and long-term stable systems. We therefore pursue a simpler binary classifier here and defer a regression algorithm to future work.

Any binary stability classification must specify a timescale, and we choose a value of 10^9 orbits. Exoplanet systems may indeed continually undergo instabilities and restructurings over their lifetimes (e.g., refs. 30 and 51). In such a situation, the quick removal of short-lived configurations naturally leaves systems with instability times comparable to their ages (42). In that case, requiring stability over 10^{11} orbits in a comparably aged system could potentially throw out the true orbital configuration. However, one could still reasonably reject configurations that destabilize within $< 10^9$ orbits, given the low likelihood of finding a system within $< 1\%$ of its lifetime of going unstable. Following previous authors (28, 31), we define instability as when a pair of planets start crossing their sphere of gravitational influence (*Materials and Methods*).

For the remainder of the paper, we refer to configurations that survive 10^9 orbits as stable and ones that do not as unstable.

Training SPOCK. We frame our task as a supervised machine-learning problem. We begin by generating a large suite of $\sim$100,000 initial conditions and perform the computationally expensive numerical integrations over 10^9 orbits to empirically label each example as stable or unstable (*Training Set*). We take 80% of these examples as a training set for our classifier and use the remaining 20% as a holdout set to test for potential overfitting with examples that were never encountered during training (*Training SPOCK*).

The input to SPOCK is then a complete initial orbital configuration: stellar and planetary masses, along with six orbital elements or positions and velocities for each of the planets. Our strategy for making a stability prediction is to first run a computationally inexpensive integration of only 10^4 orbits and, from this short snippet, numerically measure dynamically informative quantities (*Machine-Learning Model*). Given that the machine-learning model evaluation is effectively instantaneous, this represents a speed-up factor of up to 10^5. This feature engineering step allows us to take a high-dimensional set of inputs and reduce it to 10 features that more compactly encode our partial understanding of the dynamics. We then train a machine-learning classifier to take this set of summary features as input to predict the probability that the system is stable over 10^9 orbits. This is illustrated in Fig. 1.

Following a previous proof of concept (52), we use the gradient-boosted decision tree algorithm XGBoost (53). We found it significantly outperformed simple random forest and support-vector machine implementations. However, the specific choice of XGBoost was not critical. In an early comparison, we found similar results training a deep neural network (multilayer perceptron) on the same features (see also ref. 54 for an application to circumbinary planets). The most important factor for performance was the adopted set of summary metrics.

Going beyond exploratory machine-learning models (52) to a robust classifier applicable to observed compact systems required several qualitative changes. First, we relax their assumption of equal-mass planets, and we extend their horizon of stability prediction (10^7 orbits) by a factor of 100 to a relevant timescale for real systems. Second, previous work (18, 19, 22) suggests that instabilities in compact systems are driven by the overlap of MMRs. Rather than sampling phase space uniformly as done in ref. 52, we therefore choose to generate our training dataset of three-planet systems in and near such MMRs. This fills in our training sample at the locations in phase space where the dynamical behavior is changing most rapidly, and we suspect this identification of the dominant dynamics is partially responsible for the excellent generalization to more general systems presented in *Training Set* and *Machine-Learning Model*. Fig. 2 shows our training set, plotting the period ratio between the inner two planets against that of the outer two planets.

Finally, the sharp changes in dynamics at each resonance (narrow lines visible in Fig. 2) make predictions challenging in the space of traditional orbital elements (orbital periods, eccentricities, etc.). Indeed, we found that the model of ref. 52 performed poorly near MMRs. In this study, we exploit analytical transformations (55, 56) that isolate the effect of a particular resonance between a single pair of planets (*Materials and Methods*). This allows us both to effectively sample the appropriate ranges in period ratios and other orbital elements (like eccentricities) not visible in the projection of Fig. 2 and to calculate several summary features in this smoother transformed space that makes predictions simpler than in the original sharply punctuated space of Fig. 2 (*Machine-Learning Model*).

Training Set. The two-planet case is analytically solvable (18–20), while for 3+ planet systems, there is a qualitative change toward a continuous range of instability times over wider interplanetary separations (21). We posit that instabilities driven by MMR overlap in higher-multiplicity systems can be approximated by considering only adjacent planet trios. Our training set thus consists only of compact three-planet systems, and we later test the trained model's generalization to higher-multiplicity systems (*Generalization to Higher-Multiplicity Systems*). This is an

[‡]Computation timescales linearly with the number of orbits and requires $\sim$10 s per million orbits with optimized algorithms (50) and current hardware.

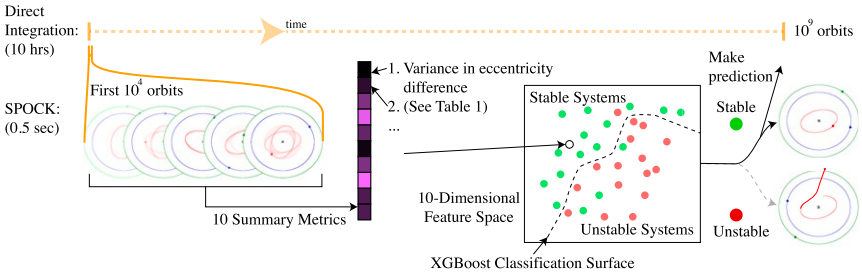

Fig. 1. A schematic illustrating how SPOCK classifies stability of a given initial configuration. The traditional approach numerically integrates the system for 10^9 orbits, requiring roughly 10 h with current hardware. SPOCK runs a much shorter 10^4 orbit integration and, from it, generates a set of 10 summary features (Table 1). These map to a point (white circle) in a 10-dimensional space, in which we have trained an XGBoost model to classify stability. SPOCK outputs an estimated probability that the given system is stable over 10^9 orbits, up to 10^5 times faster than direct integration.

important empirical test since, if true, it implies a robustness of our stability classifications to distant unseen planets. This is crucial for reliable stability constrained characterization of exoplanet systems and is consistent with previous numerical experiments with equal-separation planets showing an insensitivity to additional bodies beyond somewhat larger multiplicities of five (21), as well as theoretical arguments showing that the Fourier amplitudes of the perturbation potential due to an additional planet fall off exponentially with separation (22, 35).

To enable application to real systems, we sample unequal-mass, unequal-separation, mutually inclined, eccentric initial orbital three-planet configurations for our training set, drawing from parameter ranges typically encountered in the current multiplanet sample.

In particular, the vast majority of 3+ planet systems have been discovered by the Kepler and K2 missions as the bodies pass in front of (transit) their host star.

This implies that nearly all such systems must be approximately coplanar; otherwise, the planets would not all cross in front of the star from our vantage point (e.g., ref. 57). We therefore sample inclinations (log-uniformly and independently) from a narrow range of $[10^{-3}, 10^{-1}]$ radians (where the upper limit has been extended somewhat beyond the mutual inclinations typically inferred to also allow the modeling of additional [unobserved] nontransiting planets). The azimuthal orientations of the orbital planes (i.e., the longitudes of the ascending nodes) were drawn uniformly from $[0, 2\pi]$. This corresponds to maximum mutual-orbital inclinations of $\sim 11^\circ$.

Most planets ($\sim$85%) in the current sample of compact 3+ planet systems, where stability constraints are most informative, are smaller than Neptune. We therefore choose to independently and log-uniformly sample mass ratios to the central star from 10^{-4} (approximately two times that of Neptune to the Sun) down below the typical threshold of detectability to 10^{-7} (approximately one-third that of Mars to the Sun).

Any measure of dynamical compactness must incorporate these planetary masses. This is often expressed in terms of the separations between adjacent planets in units of their mutual Hill radius (Eq. **1**). We always initialize the innermost planet's semimajor axis at unity (since, as mentioned above, we work in units of the innermost semimajor axis) and choose to sample the separations between adjacent planets in the range from [0,30] R_H. This encompasses $\sim$80% of the currently known planets in well characterized 3+ planet systems (58). For scale, 30 R_H also roughly corresponds to the wider dynamical separation between the terrestrial planets in our solar system.

In particular, we randomly choose a planet pair (inner, outer, or nonadjacent) and randomly sample their remaining orbital parameters in or near a randomly chosen MMR within 30 R_H, as described in detail in *Materials and Methods*. Finally, we draw the remaining planet's separation from its nearest neighbor uniformly in the range [0,30] R_H. This gives rise to the extended lines in Fig. 2. Two of the planets are initialized at a particular resonant ratio (e.g., 3/2 on the x axis), while the third planet's period can span a continuous range across different configurations and is not necessarily strongly influenced by MMRs.

Orbital eccentricities and phases for the resonant pair are described in *Materials* and *Methods*, while the third planet's orbital eccentricity is drawn log-uniformly between the characteristic eccentricities imparted when inner planets overtake their outer neighbors (approximated as the ratio of the interplanetary forces to the central force from the star) and the nominal value at which adjacent orbits would cross

$$e_{\rm cross} = (a_{i+1} - a_i)/a_{i+1}\ . \qquad [2]$$

Pericenter orientations and phases along the orbit for the remaining planet are drawn uniformly from $[0, 2\pi]$. Finally, we

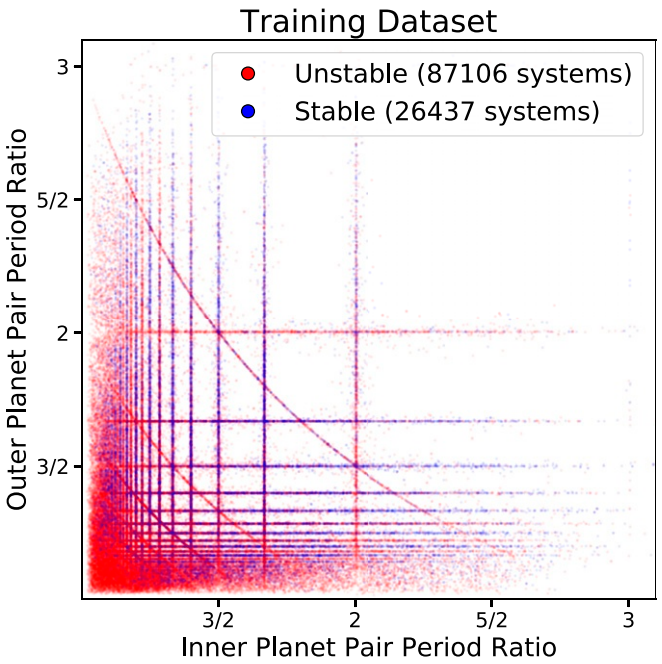

Fig. 2. Training dataset of three-planet systems in and near MMRs, which we posit drive instabilities in most compact multiplanet systems (18, 22, 31). These cluster in lines at integer period ratios between adjacent planets. The lifetimes of configurations within each of those lines can vary drastically depending on additional parameters (masses, eccentricities, phases) not visible in this projection, causing stable (blue) and unstable (red) systems to overlap.

reject any configurations that destabilize within 10^4 orbits. We refer to this dataset of 113,543 systems as the "resonant dataset."

Machine-Learning Model. We have a choice of what set of features about each system to pass the XGBoost classifier. For example, one could fully characterize a system by passing each planet's mass ratio with the star and its initial Cartesian positions and velocities. However, presumably it would be easier for the algorithm if one transformed those initial conditions to more physical parameters, like the orbital elements (semimajor axis, eccentricity, etc.). We instead choose to run a computationally inexpensive integration over 10^4 orbits and numerically measure 10 dynamically relevant quantities at 80 equally spaced outputs.

We experimented with different lengths of integrations and number of outputs. Given an integration time series, our implementation takes a few tenths of a second to compute the summary features and evaluate the XGBoost model. This is about the time required to run 10^4 orbits through N-body; so given this fixed overhead, there is no computational gain in running a shorter integration. We found that the performance gain from longer integrations was marginal, but if the feature and model evaluations were optimized (e.g., ported to C), a more careful optimization of these parameters could be valuable.

For our first two features, we store the median of the MEGNO chaos indicator (49) (over the last 10% of the short integration) and its SD (over the last 80% of the time series to avoid initial transients) as MEGNO and MEGNOstd, respectively. We also record the initial values EMcross of $e_{\rm cross}$ (Eq. **2**) for each adjacent planet pair, and we use the smallest value to identify one adjacent pair of planets as "near" and the other as "far" for these and all remaining features.

The remaining six summary statistics capture the resonant dynamics. In particular, in the near-resonant two-planet limit, only one particular combination of the eccentricities (approximately their vector difference; *Materials and Methods*) matters (55, 56),

$$\mathbf{e}_- \equiv \mathbf{e}_{i+1} - \mathbf{e}_i, \qquad [3]$$

where $\mathbf{e}_i$ is a vector pointing toward the ith planet's orbital pericenter, with a magnitude given by the orbital eccentricity.

Two of our summary features (one for each adjacent planet pair) are the SD of $|\mathbf{e}_-|$ over the timespan of the short 10^4 orbit integration, which we normalize through Eq. **2** to the value required for that planet pair to cross. Qualitatively, this can help the classifier differentiate between configurations that oscillate close to a resonant equilibrium (small variations) and are dynamically protected by the MMR, versus configurations far from equilibrium where the MMR induces large-amplitude, often destabilizing, variations.

For each adjacent planet pair, we also search for the strongest j:j-k MMR within 3% of the pair's period ratio and record its nondimensionalized strength,

$$s = \sqrt{\frac{m_i + m_{i+1}}{M_\star}} \frac{(e_-/e_{\rm cross})^{k/2}}{(jn_{i+1} - (j-k)n_i)/n_i}, \qquad [4]$$

where the m_i are the planet masses, and the n_i are the orbital mean motions ($n_i = 2\pi/P_i$ with P_i as the orbital periods). This is the appropriate expression when linearizing the dynamics, omitting a period ratio-dependent prefactor that is comparable for all of the nearby resonances (56). It is thus adequate for identifying the strongest nearby MMR, and we store its median value over the short integration as MMRstrengthnear and MMRstrengthfar.

Finally, we record the SD of $|\mathbf{e}_+|$, a complementary combination of eccentricities to $\mathbf{e}_-$ that is approximately conserved (55, 56) in the single resonance, two-planet model (*Materials* and *Methods*). Providing SPOCK with the variation of this putatively conserved $|\mathbf{e}_+|$ variable quantifies the validity of our simple analytic model. In particular, the analytical transformation is useful along isolated lines in Fig. 2, where a single resonance dominates the dynamics. The transformation breaks down (and $|\mathbf{e}_+|$ can vary significantly) at locations where resonances cross and more than one resonance drives the dynamics, as well as in the blank space between resonances in Fig. 2. However, these are typically also the easier regions to classify. Line crossings are regions where resonances are typically strongly overlapped to drive rapid chaos (59), and the dynamics vary more smoothly in the regions between strong resonances. The complementarity and flexibility of these 10 features allow SPOCK to reliably classify stability in a broad range of compact configurations.

We calculate these 10 features (summarized in Table 1) for all initial conditions in our resonant dataset and then use them to train a gradient-boosted decision tree XGBoost model (53). We adopt an 80 to 20% train-test split, performing fivefold cross-validation on the training set. We optimized hyperparameters to maximize the area under the receiver operator characteristic (ROC) curve (Fig. 3) using the hyperopt package (60). We provide our final hyperparameter values and ranges sampled in a jupyter notebook in the accompanying repository, which trains the model.

In Table 1, we also list the relative feature importances in our final model, which measure the occurrence frequency of different features in the model's various decision trees. All provide comparable information, partially by construction. We started with a much wider set of 60 features, iteratively removing less important ones. This marginally decreased the performance of our final classifier, but this is compensated by the improved interpretability of our simplified feature set. While the feature importances are close enough that one should not overinterpret their relative values, it is clear that the resonant features are providing important dynamical information.

Results

Holdout Set Performance. The accuracy of any classifier depends on the dataset. For example, it would be much harder to determine stability over 10^9 orbits on a set of configurations right at the boundary of stability, which all went unstable between 10^8 and 10^{10} orbits, than on a dataset of configurations that either go unstable within 10^3 or survived beyond 10^{15} orbits. Thus, to avoid any straw man comparisons to previous work, we follow a parallel process of training an XGBoost model using the quantities (features) considered by previous authors. This allows each model to optimize its thresholds for the training set at hand, providing a fair comparison. In particular, for "N-body," we ask the XGBoost model to predict stability based on the instability time

Table 1. Summary features in our trained model, ranked by their relative importance

Feature name	Description	Importance
EMcrossnear	Initial orbit-crossing e_- value	6,844
MMRstrengthnear	Median strength of nearest MMR	6,568
MMRstrengthfar	Median strength of nearest MMR	6,392
EPstdnear	SD of e_+ mode	6,161
EMfracstdfar	SD of e_- mode/EMcross	5,815
EMfracstdnear	SD of e_- mode/EMcross	5,509
EMcrossfar	Initial orbit-crossing e_- value	5,077
EPstdfar	SD of e_+ mode	5,009
MEGNOstd	SD of chaos indicator	4,763
MEGNO	Chaos indicator	4,350

See Machine-Learning Model for discussion. The smallest value of EMcross is used to label one adjacent pair of planets as "near" and the other as "far."

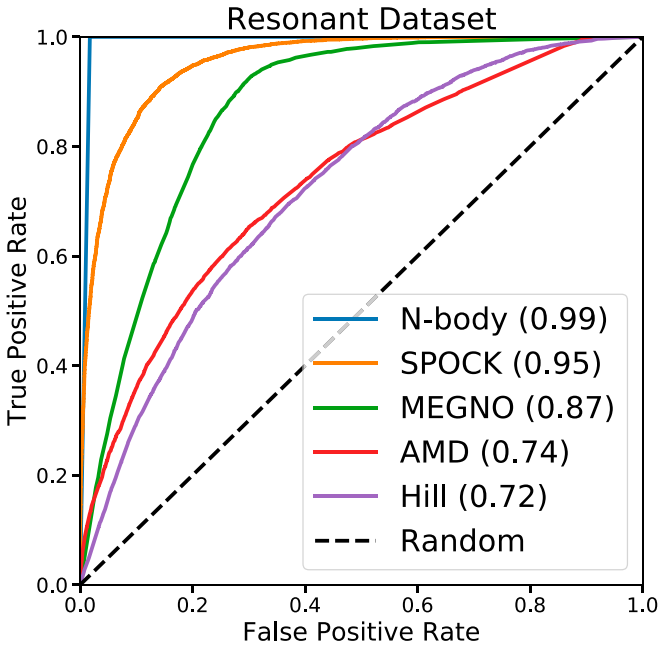

Fig. 3. Comparison of the performance of SPOCK against previous models on a holdout test set from our training data (*Materials* and *Methods*). Plots TPR (fraction of stable systems correctly classified) vs. FPR (fraction of unstable systems misclassified as stable). All models can trade off TPR vs. FPR by adjusting their threshold for how conservative to be before labeling a system as stable. The area under the curve (AUC) for each model is listed in the legend in parentheses. A perfect model would have a value of unity, and random guessing (dashed black line) would have AUC = 0.5. The blue N-body curve gives an empirical estimate of the best achievable performance on this dataset. At an FPR of 10%, SPOCK correctly classifies 85% of stable systems; MEGNO, 49%; AMD, 36%; and Hill, 30%. For a discussion of the various models, see *Previous Models*.

measured in the shadow integration (*N-Body*). For "MEGNO," we measure the chaos indicator over the same short 10^4 orbit integration as our model and pass this as a single feature to a separate XGBoost model. For "AMD," we train on two features: the system's total AMD as a fraction of the critical AMD needed for a collision between each pair of adjacent planets (44). Finally, for "Hill," we train an XGBoost model on the separations between adjacent planets as a fraction of their mutual Hill sphere radius (Eq. **1**).

Using a holdout test set of ∼20,000 integrations, we present in Fig. 3 the various models' ROC curves. ROC curves plot the classifiers' true-positive rate (TPR) (the fraction of stable systems correctly identified) vs. the false-positive rate (FPR) (the fraction of unstable systems incorrectly labeled as stable). Each model returns an estimated probability of stability and can trade off TPR vs. FPR by adjusting its threshold for how conservative it is before it labels a system as stable. A perfect model would lie in the top left corner, and random guessing would follow the dashed black line.

Different applications will have different requirements. For example, stability considerations could be important in inferring the underlying multiplicity distribution of planets in multiplanet systems detected using short time baselines of observation (e.g., with the Transiting Exoplanet Survey Satellite [TESS] mission). In that limit, it becomes important to account for the fact that it becomes harder to stably fit higher-multiplicity systems into a short observing window. Such studies estimate the underlying planet occurrences by correcting for various observation biases, typically injecting fake planetary systems into data to model the detection efficiency across the parameter range.

Injecting self-consistent, stable multiplanet configurations requires a low FPR. If a system is unstable, one wants to be confident that it will be labeled as unstable and thrown out of the analysis. If one decided that a 10% false-positive was acceptable, one could read off the corresponding TPR from Fig. 3. N-body would correctly label all stable systems, while SPOCK correctly identifies 85%. MEGNO, AMD, and Hill are not competitive, with TPR values $\leq 50\%$. MEGNO and SPOCK are roughly a factor of 10^5 times faster than N-body, while AMD and Hill sphere-separation models are effectively instantaneous since they are calculated directly from the initial conditions.

It is important to note that this is an unusually demanding test dataset, asking models to make predictions at sharp resonances where the dynamical behavior changes drastically with small changes in parameters (Fig. 2). In reality, our solar system and most exoplanet systems are not close to such MMRs (57), so one should expect the performance on typical systems to be better for all models than what is shown in Fig. 3. This approach of focusing on the most problematic resonant systems differs from the more uniform phase-space coverage used in previous work and, we expect, should yield more robust, generalizable models with fewer training examples. Conversely, the generalization of such a model trained at sharp resonances to the remaining phase space is a strong test of whether MMRs are indeed dominantly responsible for instabilities in compact planetary systems (*Generalization to Uniformly Distributed Systems*).

We now consider why previous models performed poorly. First, while the Hill sphere separations are demonstrably important quantities (21, 28, 31), they do not carry any information on other important parameters like the orbital eccentricities. One therefore should not expect a simple two-parameter classifier to yield accurate predictions, particularly near resonances where the behavior depends sensitively on combinations of several different orbital elements.

Second, AMD stability has been shown to be useful in compact two-planet systems (44, 47) and can be related to the analytical Hill stability limit in such systems (61). While it still retains important dynamical information in the 3+ planet case, we see that by itself it is a poor discriminant of stability. The most obvious problem given our MMR dataset is that AMD stability applies in the secular limit, where the effects of MMRs are ignored. As refs. 44 and 56 argue, while MMRs alter the AMD, they tend to induce oscillations that average out over a resonant cycle. However, this is only true for an isolated MMR; once several resonances overlap and motions become chaotic, AMD is not necessarily conserved. While this is not a concern for two-planet systems in the AMD-stable region (61), our integrations show empirically that there are many opportunities for MMR overlap in compact systems with three or more planets, and AMD stability is no longer a stringent criterion.

One might argue that this is asking more from AMD stability than it offers, given that it is supposed to be a worst-case scenario estimate. It only guarantees stability if the total AMD is below the value needed for collisions. Above the critical AMD, collisions are possible, but AMD stability makes no prediction one way or another. However, even if we only consider the ∼19% of systems in our resonant test set that AMD guarantees are stable, only ∼49% actually are.

Finally, for the MEGNO model, a small fraction (∼2%) of the systems that it found to be chaotic (taken as a value of MEGNO after 10^4 orbits > 2.5) are nevertheless stable. Even if a system is chaotic (i.e., nearby initial conditions diverge exponentially), it still needs enough time for the eccentricities to diffuse to orbit-crossing values. For example, the GJ876 system has a Lyapunov (chaotic) timescale of only about 7 y, despite the system being of order a billion years old (62). Determining that an orbit is chaotic is therefore strongly informative but not sufficient to determine long-term stability. More problematically, 55% of the systems

with MEGNO values consistent with being regular (nonchaotic) were, in fact, unstable. This is because MEGNO can only measure chaos on the timescale of the short integration, so systems with Lyapunov times longer than 10^4 orbits can nevertheless go unstable and be misclassified by MEGNO alone. In summary, determining that an orbit is chaotic with MEGNO in a short integration is typically a reliable indicator that the system is not long-term stable, but a MEGNO value consistent with a regular orbit is not a robust measure of long-term stability.

By combining MEGNO with features capturing the MMR dynamics, SPOCK substantially improves on these previous models.

Generalization to Uniformly Distributed Systems. An important concern with machine-learning models is whether they will generalize beyond the training dataset. Since there are no theoretical generalization bounds for modern techniques like decision trees and neural networks, measuring generalization to a holdout set and out-of-distribution data are essential. In particular, have they learned something meaningful about the underlying physics, or have they simply memorized the particulars of the training set? We perform two empirical tests.

First, we generate a complementary dataset of 25,000 uniformly sampled configurations, spanning a continuous range in period ratios, and not necessarily close to MMRs. This is more representative of typical exoplanet systems that have been discovered, with one important difference. We choose our sampling ranges to yield roughly a comparable number of stable and unstable configurations ($\sim$40% were stable), while observed systems are naturally biased toward stable regions of phase space since unstable configurations are short-lived.

The procedure and parameter ranges are the same as in our training set, except we now sample all planets' orbital elements like we did the third planet above (separations uniform from [0,30] R_H, eccentricities log-uniform up to the value at which orbits cross, and all phases uniformly from $[0, 2\pi]$). We plot the projection of this "random dataset" into the space spanned by the Hill-radius separations between adjacent planets in Fig. 4, *Top*, where they uniformly fill the plane.

It is easier to predict stability on this dataset for at least two important reasons. First, most configurations are not particularly close to strong MMRs where the dynamical behavior changes sharply. Second, while in the resonant training dataset, we restricted ourselves to systems that survived longer than our short integrations of 10^4 orbits, in reality, many unstable configurations will be extremely short-lived. In our random dataset, we therefore allow for any instability time, which is more representative of typical applications. This will in particular significantly boost the performance of both the SPOCK and MEGNO models, since they will be able to confidently classify the configurations that go unstable within the span of their short integrations.

We plot the performance of all models (trained on the resonant dataset; Fig. 2) on our random dataset in Fig. 4. All models improve as expected, particularly SPOCK and MEGNO. At an FPR of 10%, N-body correctly classifies 99.8% of systems; SPOCK, 94%; MEGNO, 87%; AMD, 74%; and Hill, 39%. Over the range of FPRs in Fig. 4, SPOCK correctly labels approximately half of the systems misclassified by MEGNO.

The fact that our SPOCK classifier, trained on a discrete set of near-resonant systems (Fig. 2), performs strongly on this uniform dataset supports our assertion that instabilities in compact multiplanet systems are dominantly driven by MMRs. If instead we let SPOCK train on 80% of our random dataset and test on the remaining 20%, the TPR quoted above only rises by $\sim$2%, suggesting our model can robustly classify a wide range of compact three-planet systems.

Generalization to Higher-Multiplicity Systems. Influenced by previous work (21), we hypothesized that the simplest case for understanding instabilities within 10^9 orbits in multiplanet systems is that of three planets. A natural question is therefore how well our model, trained on three-planet systems, generalizes to higher-multiplicity cases.

We test our model's generalization on previously published numerical integrations of five equal-mass planets on coplanar, equally spaced, and initially circular orbits (31). This is in stark contrast to our systems of three unequal-mass planets on mutually inclined, unevenly spaced, and eccentric orbits. Indeed, the integrations only varied the separation between adjacent planets, corresponding to a diagonal line from the bottom left to the top right of Fig. 2. This passes through the many intersections

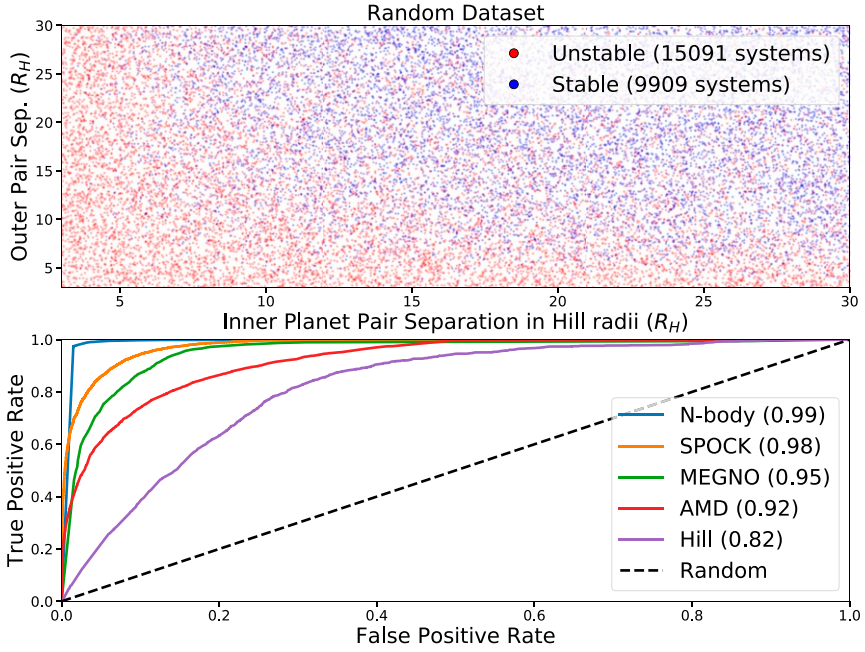

Fig. 4. *Top* shows the projection of our "random" dataset of 25,000 configurations, uniformly sampling the separation between adjacent planets. This is more representative of discovered compact exoplanet systems than the resonant training dataset. *Bottom* is analogous to Fig. 3. At an FPR of 10%, SPOCK correctly classifies 94.2% of stable systems, despite having been trained on a different dataset of near-resonant configurations (Fig. 2).

between vertical and horizontal MMR lines, where our analytic transformations (assuming the influence of only a single MMR) are most problematic. All other parameters were fixed, so there are very few examples in our much higher-dimensional training set that fall near this one-dimensional line, rendering it a particularly stringent test of our model. In particular, if SPOCK had simply memorized the particulars of our training set, it should not be able to effectively predict on systems drawn from a very different configuration distribution.

As a simple prescription for predicting stability on 4+ planet systems using SPOCK, we feed all adjacent trios to our three-planet classifier and retain the lowest stability probability. This is a simplification but should provide a reasonable approximation in cases where instabilities are driven by perturbations on an MMR between a particular pair of planets. We argue this typically is the case in compact systems.

Fig. 5, *Top* shows the instability time recorded by the 17,500 N-body integrations of ref. 31, plotted against the separation between adjacent planets (normalized by their mutual Hill radius [Eq. **1**]). As above, we color-code systems that went unstable within 10^9 orbits as red and stable systems as blue. While, above, we considered binary classification as stable or unstable, in Fig. 5, *Bottom*, we now plot the probability of stability estimated by SPOCK for each of the initial conditions in order to better visualize the model's sensitivity to the structure in instability times visible in Fig. 5, *Top* (each point along any of the SPOCK ROC curves above corresponds to the TPR and FPR obtained when setting a particular stability probability threshold).

We see that SPOCK gives reliable results, despite having been trained on very different configurations of resonant and near-resonant configurations of fewer planets. Fig. 5, *Top* also shows that SPOCK recognizes each of the dips in instability times, which correspond to the locations of MMRs (31), and adjusts its stability probability accordingly. Misclassifications are largely limited to the boundaries between stable and unstable configurations in Fig. 5, *Top*. We note that near this boundary, classification is ambiguous—some of these systems would also be "misclassified" by direct N-body integrations. Using the same threshold as in Holdout Set Performance (chosen to yield an FPR on our resonant holdout set of 10%), the TPR across this test set of 17,500 integrations is 94%, with an FPR of 6%. The fact that our model trained on three-planet systems generalizes to higher multiplicities supports our assertion at the outset that planet trios are prototypical cases that can be extended to higher numbers of planets.

An Application. As an example, we considered the characterization of four observed compact three-planet systems (Kepler-431, Kepler-446, EPIC-2108975, and LP-358-499), none of which is near strong MMRs. We again focus on compact systems, since we should be able to reject a larger range of masses and orbital eccentricities for these more delicate configurations. All four systems gave similar results, so we focus on Kepler-431, a system of three transiting Earth-sized planets (which gave the second-worst performance).

The planetary transits across the host star strongly constrain the planets' orbital periods and physical sizes. The masses and especially the orbital eccentricities remain highly uncertain. As a simple exercise, we sample the planetary masses from a mass–radius relationship (63) and sample eccentricities log-uniformly between $[10^{-4}, 0.18]$ for each of the three planets independently (with the upper limit representing the value at which the inner two orbits would cross). Since these are transiting planets, we draw inclinations from an edge-on configuration uniformly from 10^{-3} radians to the angular size of the star as seen from the planet, $R_\star/a_i$ (with $R_\star$ the stellar radius and a_i the ith planet's semimajor axis). All remaining angles are drawn uniformly from $[0, 2\pi]$, and we assume a stellar mass of 1.07 solar masses. We draw 1,500 configurations in this way, and for each one, run both direct N-body integrations and our SPOCK classifier.

Adopting the same stability probability threshold from Holdout Set Performance, we obtain the results plotted in Fig. 6. To visualize the phase space, in the top row of Fig. 6, we provide polar plots of the middle planet's eccentricity vector (with the distance from the origin giving the eccentricity and the polar angle the direction toward pericenter). Fig. 6, *Top Left* color codes stable and unstable configurations obtained through direct N-body. Fig. 6, *Top Right* shows the predictions from SPOCK, yielding an FPR of 9% and TPR of 97%.

While the expected trend of instability toward high eccentricities is born out in the top row of Fig. 6, many unstable configurations remain near the origin at zero eccentricity due to other system parameters not visible in this projection. However, by developing a classifier with a comparatively small number of physically motivated features, we can gain insight into the stability constraints by projecting the configurations onto the transformed resonant space used by the model. In the bottom row of Fig. 6, we consider the eccentricity modes e_- (Eq. **3**) that dominate the MMR dynamics between each adjacent pair of planets (first and second planet on the x axis; second and third on the y axis). We see that our feature space incorporating our analytical understanding of the resonant dynamics much more cleanly separates the stable and unstable systems, even in this two-dimensional projection. This both visually shows how our engineered features help the algorithm's performance and clarifies the particular combinations of parameters specifically constrained by stability.

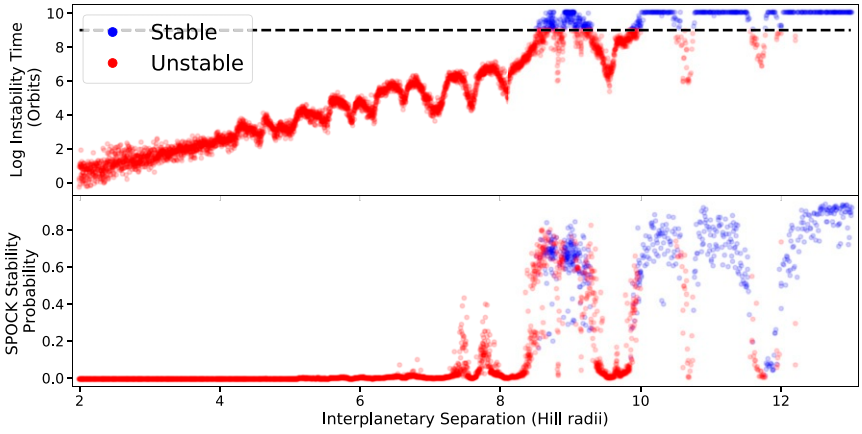

Fig. 5. Generalization of SPOCK (trained on eccentric, inclined, unequally spaced, unequal-mass three-planet systems in and near MMRs) to integrations from ref. 31 of initially circular, coplanar, equally spaced, equal-mass five-planet systems. The separation between all adjacent pairs of planets increases along the *x* axis. *Top* shows the instability times measured by direct integration, with dips corresponding to MMRs (31). *Bottom* shows the stability probability predicted by SPOCK. Taking the probability of stability threshold of 0.34 used in Holdout Set Performance, the true-positive rate is 94% and the false-positive rate 6% on this dataset.

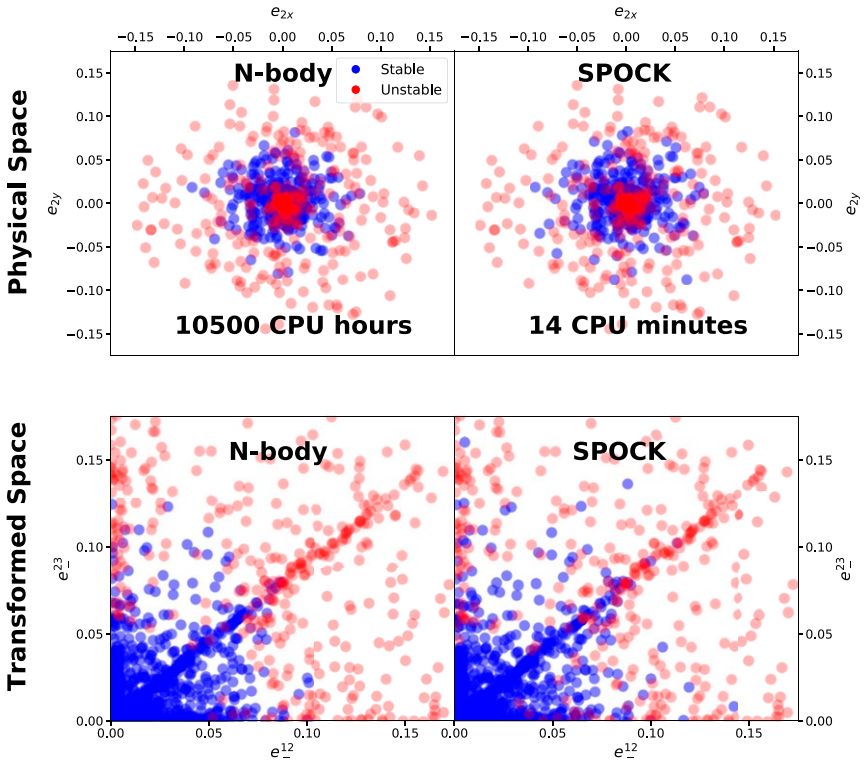

Fig. 6. *Top Left* and *Top Right* are polar plots of the middle planet's eccentricity vector in the Kepler 431 compact three-planet system (distance from origin is the eccentricity; polar angle, the direction toward pericenter). *Top Left* color codes 1,500 stable and unstable configurations through direct integration, while *Top Right* shows the stability predictions from SPOCK—96% of points agree. The bottom row projects the configurations into the transformed space used by our model features (see *Machine-Learning Model*). The stability boundary separates much more cleanly in this space that incorporates our understanding of the resonant dynamics, showing visually why our engineered features help the algorithm's performance.

Finally, we note that in this closely packed system, stability is indeed constraining. We constrain the free eccentricities[§] of the inner and outer pair of planets to be below 0.051 and 0.053, respectively (84th percentile limit). Such eccentricity limits, which constrain the degree of dynamical excitation in the system's past (64, 65), are significantly stronger than those inferred from radial velocity (e.g., ref. 66) or transit duration measurements (67, 68) for such low-mass planets, which dominate the population (e.g., ref. 69). Within a factor of a few, this approaches the exquisite constraints achievable by modeling transit-timing variations (TTVs), which are typically only measurable when planets are close to strong MMRs, with accurate photometry, and with long observation baselines (70). In particular, TTVs are not detected in any of the four Kepler systems we considered. TTV modeling has been an extremely productive method with the long observation baselines of the Kepler mission (e.g., refs. 14 and 16). However, the much shorter observing windows of Kepler's successor, the TESS, implies that only $\sim$10 planets are expected to be constrained by TTVs (71) during its prime mission. This places stability constrained characterization as a powerful complementary method for understanding multiplanet systems.

Limits. Finally, we present an instructive case where SPOCK fails, for systems constrained by above-mentioned TTVs. Transiting planets that do not interact with one another would pass in front of their host stars like perfect clocks with a constant orbital period. However, their mutual gravitational tugs can cause transit times to periodically pull ahead and fall behind. This is a particularly strong effect near MMRs, which induce sinusoidal TTVs (70, 72).

We considered six systems that exhibit TTVs and, in particular, the three-planet Kepler-307 system (73) (outermost planet only a candidate). In all cases, the transit times have been fit to infer planet masses and orbital parameters with Markov chain Monte Carlo (MCMC). We choose to sample 1,500 configurations from the resulting posterior, and again run N-body integrations to compare with SPOCK predictions as in An Application.

Interestingly, SPOCK fails on all of them. In the case of Kepler-307, the FPR is 87% (Fig. 7). An important cost to consider with complex models is the difficulty in diagnosing problems such as these when they come up. Our original SPOCK model generated 60 summary features from short integrations, and in fact slightly outperformed our final adopted model on the holdout set in Fig. 3. However, we chose to trade these marginal performance gains for the improved interpretability of our smaller set of 10 physically relevant features, and this reveals the reason for the poor performance in Fig. 7.

The inner two planets in this system are near a 5:4 MMR (period ratio $\sim$1.255), while the third planet is significantly further separated (period ratio between the outer two planets $\sim$1.79). As mentioned above, the MMR dynamics between a pair of planets are driven by a particular combination of the orbital eccentricities e_- (Eq. **3**). In this case, because the observed TTVs are driven by a 5:4 MMR between the inner two planets, the TTVs observed in the data specifically constrain this planet pair's e_- mode. If we again transform the space in the top row of Fig. 7 to that spanned by the e_- modes for both adjacent pairs like in Fig. 6, we see that the sample of configurations collapses to a thin vertical line.

The problem is therefore that while SPOCK would typically help to constrain e_- by ruling out unstable values, the MMR-driven TTVs have already allowed the MCMC fit to narrow down the e_- mode for the inner pair of planets to an exquisitely narrow range of 0.0088 ± 0.0004. Thus, samples from the MCMC posterior have already removed configurations along directions in which SPOCK has strong discerning power, leaving only points along directions that are difficult to separate from the short integrations. The "MEGNO" model similarly fails with an FPR of 57%.

[§]We quote free eccentricities typically quoted for TTV constraints $Z \approx |\mathbf{e}_-|/\sqrt{2}$ (see *Materials and Methods* for a more direct comparison).

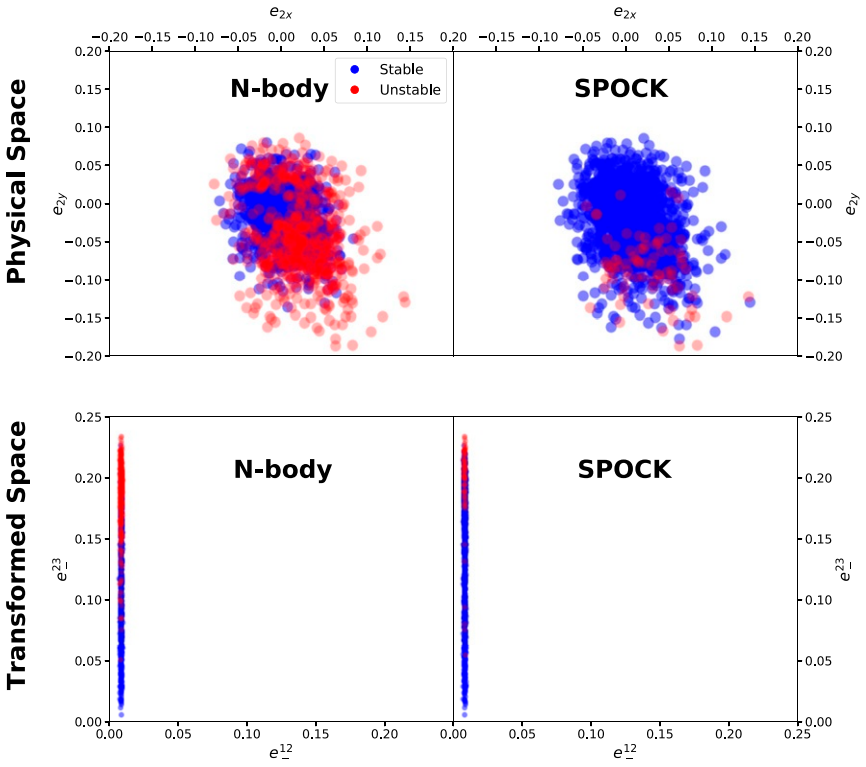

Fig. 7. Case of the three-planet Kepler-307 system, where SPOCK predictions fail. We sample configurations (points) from the posterior of a MCMC fit to the observed TTVs in the system. The top and bottom rows are analogous to Fig. 6. Transforming to the two-body resonant variables used by SPOCK in the bottom row shows the reason for the poor performance. The previous TTV fit has constrained the eccentricity mode dominantly driving the MMR dynamics of the inner two planets to an extremely narrow range. This leaves points only along a direction that does not strongly influence the short integrations we use to generate features.

If we constrained the above system blindly, across the full range of possible eccentricities, SPOCK's performance would be comparable to the results in Holdout Set Performance. In this case, however, observational data has strongly constrained the planets' resonant dynamics, leaving only configurations with very similar SPOCK features and leading to unreliable predictions. Presumably, improved models would incorporate additional features that better separate the stable and unstable configurations in Fig. 7. Such SPOCK failures should be rare, but TTV-constrained configurations are an important, instructive counterexample. One can test for such situations empirically by looking for clustering of configurations in SPOCK's feature space. An important advantage of SPOCK's physically meaningful features is that it facilitates the interpretation of any such clusterings.

Conclusion

We have presented the SPOCK, a machine-learning model capable of classifying stability of compact 3+ planet systems over 10^9 orbits. SPOCK is up to 10^5 times faster than direct N-body integration and is significantly more accurate (Figs. 3 and 4) than stability predictions using AMD stability (44), Hill-sphere separations (e.g., refs. 21, 22, and 26), or the MEGNO chaos indicator (e.g., ref. 46).

This computationally opens up the stability-constrained characterization of compact multiplanet systems, by rejecting unphysical, short-lived candidate orbital configurations. In the Kepler-431 system with three tightly packed, approximately Earth-sized planets, we constrained the free eccentricities of the inner and outer pair of planets to both be below 0.05 (84th percentile upper limits). Such limits are significantly stronger than can currently be achieved for small planets through either radial velocity or transit duration measurements and within a factor of a few from TTVs. Given that the TESS mission's typical 30-d observing windows will provide few strong TTV constraints (71), SPOCK computationally enables stability constrained characterization as a productive complementary method for extracting precise orbital parameters in compact multiplanet systems.

Our training methodology and tests also clarify the dynamics driving instabilities in compact exoplanet systems. Our model, trained solely with configurations in and near MMRs, accurately predicts instabilities within 10^9 orbits across the full phase space of typical compact systems (*Generalization to Uniformly Distributed Systems*). This is strong confirmation that rapid instabilities, on timescales much shorter than the typical $\sim 10^{-11}$-orbit ages of observed systems, are dominantly driven by the overlap of MMRs (18, 22, 31).

Instabilities can also occur on longer timescales through the overlap of secular resonances. As opposed to MMRs between planets' orbital rates, secular resonances represent commensurabilities between the much slower rates at which orbits precess. This is the case for our solar system, which has a dynamical lifetime $> 10^{10}$ orbits (39, 45, 62). SPOCK is not trained to detect such slow instabilities but self-consistently classifies the solar system as stable over 10^9 orbits.

Recent work (74) suggests that instabilities in compact systems are driven through the overlap of such secular resonances. While this may seem in tension with our focus on MMRs, this paints a self-consistent picture. Short-lived configurations eliminate themselves, rearranging and dynamically carving out the distribution of planetary systems that survive to the present day. This idea has been advanced from several perspectives (30, 51, 75–77). MMR-driven instabilities happen quickly compared with the typical ages of observed systems, leaving today only systems that destabilize through slower secular instabilities. This also clarifies that secular analyses such as AMD stability are valuable dynamical classifications for observed systems (44), despite their poor identification of short-term instabilities (*Results*).

We also showed that short-term, MMR driven instabilities are local, as expected from the lack of strong MMRs beyond period ratios of 2:1 (22, 78). In particular, we showed that our model, trained on three-planet systems, can be applied to adjacent trios of planets in higher-multiplicity systems to classify stability over

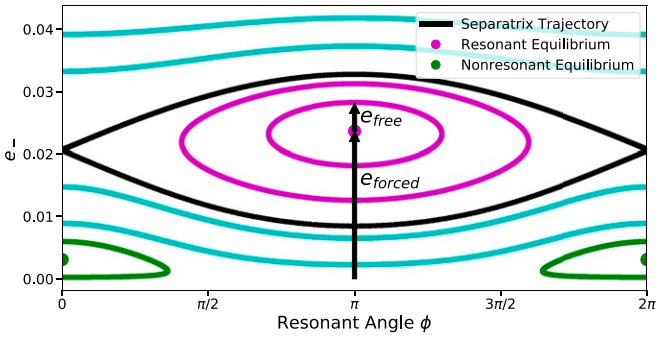

Fig. 8. Phase portrait of the dynamics near a MMR. See *Materials and Methods* for discussion.

10^9 orbits. This implies that stability constrained characterization is robust against distant, unseen planets (an important consideration for detection methods heavily biased against finding such bodies). This is not the case for longer-timescale secular instabilities. For example, exodynamicists detecting only the inner solar system would infer a much more stable system than is actually the case (e.g., refs. 39 and 62).

By identifying the dominant dynamics driving the instabilities we aimed to classify, and incorporating this directly into both the training set and the set of features used for our machine-learning model, we have trained a robust classifier of multiplanet stability over 10^9 orbits. This approach also allowed us to both test our assumptions and understand regions of phase space where SPOCK should fail. This can be a useful blueprint for exploiting the often extensive domain knowledge available in scientific applications of machine learning.

We make our ∼1.5 million CPU-hour training set publicly available (79) and provide an open-source package, documentation, and examples (https://github.com/dtamayo/spock) for efficient classification of planetary configurations that will live long and prosper.

Materials and Methods

Resonant Dataset. We initialize our near-resonant pair of planets by first identifying all of the first-order ($n:n-1$) and second-order ($n:n-2$) MMRs in the range from [3.5, 30] mutual Hill radii ([3.5, 60] mutual Hill radii for nonadjacent planets),[¶] which represent the strongest set of resonances (38). We then randomly assign the pair to one of the resonances in this list.

A pair planets near a first-order $n:n-1$ MMR, whose orbits evolve in a 12-dimensional phase space, can to excellent approximation be reduced to a 2-dimensional dynamical system through a transformation of variables involving several conserved quantities (55). This transformation has recently been generalized for $n:n-2$ and higher-order MMRs (56) (see also ref. 20).

First, at low eccentricities, the eccentricity and inclination evolution decouple. Therefore, we sample the planets' orbital inclinations and orbital plane orientations randomly as described in Generalization to Uniformly Distributed Systems. For the eccentricities, the intuition is that only one mode (or combination) matters. In particular, a combination $\mathbf{Z}_+$, approximately the "center-of-mass" eccentricity vector, is conserved, while $\mathbf{Z}$ drives the dynamics (e.g., ref. 56),

$$\sqrt{2}\mathbf{Z} \approx \mathbf{e}_2 - \mathbf{e}_1 \equiv \mathbf{e}_-, \quad \mathbf{Z}_+ \approx \frac{m_1\mathbf{e}_1 + m_2\mathbf{e}_2}{m_1+m_2} \equiv \mathbf{e}_+ \approx constant, \qquad [5]$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the planets' orbital eccentricity vectors, and m_1 and m_2 are their respective masses. The Z and Z_+ vectors incorporate additional coefficients on the eccentricity vectors that depend on the particular MMR. With the exception of the 2:1 MMR, these coefficients are within ∼10% of unity (19) and converge to one as the period ratio shrinks (56). Models trained with (Z, Z_+) exhibit similar performance to ones trained with (e_-, e_+), so we adopt the latter to avoid discontinuities in our features near discrete resonances.

Finally, one can combine the period ratio and other orbital parameters to define a metric for proximity to the resonance. The closer the system is to the resonance, the higher $\mathbf{e}_-$ will be forced by the MMR. We sample the relevant range in proximities to resonance by drawing the eccentricity forced by the resonance e_{forced} (Fig. 8) log-uniformly from the minimum value (predicted by the analytic theory to yield an isolated resonant island) (56) to the nominal orbit-crossing value e_{cross} (Eq. **2**). Choosing e_{forced} defines a family of different possible trajectories, depending on the remaining initial conditions, which we plot in Fig. 8. The variable on the x axis is the resonant angle ϕ, which is the particular combination of the pericenter orientations and phases corresponding to the given MMR—physically, it tracks the position at which conjunctions occur, where planets most strongly interact. The choice of e_{forced}, or proximity to resonance, moves the entire resonant island (bounded by the black curve) up and down.

The final important parameter is then the free eccentricity e_{free}, which measures the distance from the resonant equilibrium (magenta dot near the center of Fig. 8). A system initialized with $e_{\text{free}} = 0$ at the center of the island will remain at the equilibrium. Configurations initialized with small nonzero values of e_{free} will oscillate (librate) around the equilibrium like a pendulum. There is a critical value of e_{free} beyond which the system will no longer librate around the equilibrium but instead circulates. The boundary trajectory, which forms a "cat's eye" splitting these two types of behaviors, is called the separatrix (black curve in Fig. 8).

To fill in this range of behaviors, we sample e_{free} log-uniformly from $[3\times10^{-3}, 3]$ of the distance to the separatrix along $\phi=\pi$. In this way, the resonant pair of planets spans the range from being in resonance to being outside the resonant region, but still having their dynamics strongly influenced by the MMR. Many of the planets discovered by the Kepler mission that exhibit TTVs lie in the region near (but outside) strong MMRs (e.g., ref. 81). When the drawn e_{free} is large enough to move beyond the bottom of the plot, we wrap around and initialize the system at $\phi=0$. This allows us also to sample the other island shown in green in Fig. 8, which has an equilibrium at small values of e_-. Once the initial conditions are allowed to evolve, they fill the space in Fig. 8, as shown by the few plotted sample trajectories. For the conserved quantity e_+, we sample it log-uniformly within the same range as e_{forced}. We draw the remaining nuisance angles uniformly from $[0, 2\pi]$.

We initialized the resonant pair of planets using the open-source celmech package (https://github.com/shadden/celmech), which is based on ref. 56. celmech includes an application programming interface (API) for initializing resonant orbital configurations from the above parameters, and we include the scripts and random seeds used to generate our training sets in the data deposition accompanying this paper.

Numerical Integrations. All integrations were performed with WHFast (32), part of the REBOUND N-body package (82). We adopted a timestep of ∼3.4% of the innermost planet's orbital period. If any planets' Hill spheres overlapped, the simulation was ended and the instability time recorded. The specific halting condition is not important (80), as once Hill spheres start crossing, the system becomes an orbit-crossing tangle on orbital timescales.[#]

The integrations analyzed in this work were saved in the REBOUND SimulationArchive format, which enables exact, machine-independent reproducibility of results (83). We provide instructions and the scripts necessary to reproduce the figures in the accompanying data deposition.

ACKNOWLEDGMENTS. We are grateful to Antoine Petit and the anonymous reviewers for insightful suggestions that significantly improved the manuscript. Support for D.T. was provided by NASA through NASA Hubble Fellowship Program Grant HST-HF2-51423.001-A, awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under Contract NAS5-26555. H.R. has been supported by Natural Sciences and Engineering Research Council of Canada Discovery Grant RGPIN-2014-04553 and the Center for Planetary Sciences at the University of Toronto Scarborough. D.T. is a NASA Hubble Fellowship Program Sagan Fellow.

[¶]The lower limit is the Hill stability limit (80) was chosen to avoid immediate instabilities.

[#]It might still take a long time for small planets close to their host star to find one another and collide (34); however, in the context of applying stability constraints, we are usually interested in the time to instability defined such that the system architecture becomes inconsistent with the typically observed multiplanet system of approximately planar, concentric near-circular orbits.

1. A. N. Kolmogorov, On conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl. Akad. Nauk SSSR.* **98**, 527–530 (1954).
2. J. Möser, On invariant curves of area-preserving mappings of an annulus. *Nachr. Akad. Wiss. Göttingen* **2**, 1–20 (1962).
3. V. Arnol'd, Proof of a theorem of an Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian. *Russ. Math. Surv.* **18**, 9 (1963).
4. M. Hénon, Explorationes numérique du probleme restreint IV: Masses egales, orbites non periodique. *Bullettin Astronomique* **3**, 49–66 (1966).
5. A. Celletti, L. Chierchia, Kam stability for a three-body problem of the solar system. *Zeitschrift für angewandte Mathematik und Physik ZAMP* **57**, 33–41 (2005).
6. S. M. Mills *et al.*, A resonant chain of four transiting, sub-neptune planets. *Nature* **533**, 509–512 (2016).
7. M. Gillon *et al.*, Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1. *Nature* **542**, 456–460 (2017).
8. J. H. Steffen *et al.*, Transit timing observations from Kepler: VII. confirmation of 27 planets in 13 multiplanet systems via transit timing variations and orbital stability. *Mon. Not. Roy. Astron. Soc.* **428**, 1077–1087 (2013).
9. D. Tamayo, A. H. Triaud, K. Menou, H. Rein, Dynamical stability of imaged planetary systems in formation: Application to HL tau. *Astrophys. J.* **805**, 100 (2015).
10. J. J. Wang *et al.*, Dynamical constraints on the HR 8799 planets with GPI. *Astron. J.* **156**, 192 (2018).
11. B. Quarles, E. V. Quintana, E. Lopez, J. E. Schlieder, T. Barclay, Plausible compositions of the seven TRAPPIST-1 planets using long-term dynamical simulations. *Astrophys. J. Lett.* **842**, L5 (2017).
12. D. Tamayo, H. Rein, C. Petrovich, N. Murray, Convergent migration renders TRAPPIST-1 long-lived. *Astrophys. J. Lett.* **840**, L19 (2017).
13. E. J. Rivera *et al.*, The Lick-Carnegie exoplanet survey: A Uranus-mass fourth planet for GJ 876 in an extrasolar Laplace configuration. *Astrophys. J.* **719**, 890 (2010).
14. D. Jontof-Hutter, J. J. Lissauer, J. F. Rowe, DC. Fabrycky, Kepler-79's low density planets. *Astrophys. J.* **785**, 15 (2014).
15. L. A. Buchhave *et al.*, A 1.9 Earth radius rocky planet and the discovery of a non-transiting planet in the Kepler-20 system. *Astron. J.* **152**, 160 (2016).
16. S. Hadden, Y. Lithwick, Kepler planet masses and eccentricities from TTV analysis. *Astron. J.* **154**, 5 (2017).
17. S. L. Grimm *et al.*, The nature of the TRAPPIST-1 exoplanets. *Astron. Astrophys.* **613**, A68 (2018).
18. J. Wisdom, The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body problem. *Astron. J.* **85**, 1122–1133 (1980).
19. K. M. Deck, M. Payne, M. J. Holman, First-order resonance overlap and the stability of close two-planet systems. *Astrophys. J.* **774**, 129 (2013).
20. S. Hadden, Y. Lithwick, A criterion for the onset of chaos in systems of two eccentric planets. *Astron. J.* **156**, 95 (2018).
21. J. E. Chambers, G. W. Wetherill, A. P. Boss, The stability of multi-planet systems. *Icarus* **119**, 261–268 (1996).
22. A. C. Quillen, Three-body resonance overlap in closely spaced multiple-planet systems. *Mon. Not. Roy. Astron. Soc.* **418**, 1043–1054 (2011).
23. A. C. Quillen, R. S. French, Resonant chains and three-body resonances in the closely packed inner Uranian satellite system. *Mon. Not. Roy. Astron. Soc.* **445**, 3959–3986 (2014).
24. K. Yoshinaga, E. Kokubo, J. Makino, The stability of protoplanet systems. *Icarus* **139**, 328–335 (1999).
25. F. Marzari, S. J. Weidenschilling, Eccentric extrasolar planets: The jumping Jupiter model. *Icarus* **156**, 570–579 (2002).
26. J. L. Zhou, D. N. Lin, Y. S. Sun, Post-oligarchic evolution of protoplanetary embryos and the stability of planetary systems. *Astrophys. J.* **666**, 423 (2007).
27. P. Faber, A. C. Quillen, The total number of giant planets in debris discs with central clearings. *Mon. Not. Roy. Astron. Soc.* **382**, 1823–1828 (2007).
28. A. W. Smith, J. J. Lissauer, Orbital stability of systems of closely-spaced planets. *Icarus* **201**, 381–394 (2009).
29. Y. Matsumoto, M. Nagasawa, S. Ida, The orbital stability of planets trapped in the first-order mean-motion resonances. *Icarus* **221**, 624–631 (2012).
30. B. Pu, Y. Wu, Spacing of Kepler planets: Sculpting by dynamical instability. *Astrophys. J.* **807**, 44 (2015).
31. A. Obertas, C. Van Laerhoven, D. Tamayo, The stability of tightly-packed, evenly-spaced systems of earth-mass planets orbiting a sun-like star. *Icarus* **293**, 52–58 (2017).
32. H. Rein, D. Tamayo, WHFast: A fast and unbiased implementation of a symplectic wisdom–holman integrator for long-term gravitational simulations. *Mon. Not. Roy. Astron. Soc.* **452**, 376–388 (2015).
33. N. Hussain, D. Tamayo, Fundamental limits from chaos on instability time predictions in compact planetary systems. *Mon. Not. Roy. Astron. Soc.* **491**, 5258–5267 (2019).
34. D. R. Rice, F. A. Rasio, J. H. Steffen, Survival of non-coplanar, closely packed planetary systems after a close encounter. *Mon. Not. Roy. Astron. Soc.* **481**, 2205–2212 (2018).
35. A. Yalinewich, C. Petrovich, Nekhoroshev estimates for the survival time of tightly packed planetary systems. *Astrophys. J. Lett.* **892**, L11 (2020).
36. B. Funk, G. Wuchterl, R. Schwarz, E. Pilat-Lohinger, S. Eggl, The stability of ultra-compact planetary systems. *Astron. Astrophys.* **516**, A82 (2010).
37. D. H. Wu, R. C. Zhang, J. L. Zhou, J. H. Steffen, Dynamical instability and its implications for planetary system architecture. *Mon. Not. Roy. Astron. Soc.* **484**, 1538–1548 (2019).
38. C. D. Murray, S. F. Dermott, *Solar System Dynamics* (Cambridge University Press, Cambridge, UK, 1999).
39. Y. Lithwick, Y. Wu, Theory of secular chaos and Mercury's orbit. *Astrophys. J.* **739**, 31 (2011).
40. K. Batygin, A. Morbidelli, M. J. Holman, Chaotic disintegration of the inner solar system. *Astrophys. J.* **799**, 120 (2015).
41. P. Laplace, Mémoire sur les inégalités séculaires des planetes et des satellites mem. *Acad. R. Sci. Paris* Oeuvres completes **XI**, 49–92 (1784).
42. J. Laskar, The chaotic motion of the solar system: A numerical estimate of the size of the chaotic zones. *Icarus* **88**, 266–291 (1990).
43. J. Laskar, On the spacing of planetary systems. *Phys. Rev. Lett.* **84**, 3240 (2000).
44. J. Laskar, A. Petit, AMD-stability and the classification of planetary systems. *Astron. Astrophys.* **605**, A72 (2017).
45. J. Laskar, M. Gastineau, Existence of collisional trajectories of Mercury, Mars, and Venus with the Earth. *Nature* **459**, 817–819 (2009).
46. C. Migaszewski, M. Słonina, K. Goździewski, A dynamical analysis of the Kepler-11 planetary system. *Mon. Not. Roy. Astron. Soc.* **427**, 770–789 (2012).
47. A. C. Petit, J. Laskar, G. Boué, AMD-stability in the presence of first-order mean motion resonances. *Astron. Astrophys.* **607**, A35 (2017).
48. F. Marzari, Dynamical behaviour of multiplanet systems close to their stability limit. *Mon. Not. Roy. Astron. Soc.* **442**, 1110–1116 (2014).
49. P. M. Cincotta, C. M. Giordano, C. Simó, Phase space structure of multi-dimensional systems by means of the mean exponential growth factor of nearby orbits. *Phys. Nonlinear Phenom.* **182**, 151–178 (2003).
50. J. Wisdom, M. Holman, Symplectic maps for the N-body problem. *Astron. J.* **102**, 1528–1538 (1991).
51. K. Volk, B. Gladman, Consolidating and crushing exoplanets: Did it happen here?. *Astrophys. J. Lett.* **806**, L26 (2015).
52. D. Tamayo *et al.*, A machine learns to predict the stability of tightly packed planetary systems. *Astrophys. J. Lett.* **832**, L22 (2016).
53. T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2016), pp. 785–794.
54. C. Lam, D. Kipping, A machine learns to predict the stability of circumbinary planets. *Mon. Not. Roy. Astron. Soc.* **476**, 5692–5697 (2018).
55. W. Sessin, S. Ferraz-Mello, Motion of two planets with periods commensurable in the ratio 2:1 solutions of the Hori auxiliary system. *Celestial Mech.* **32**, 307–332 (1984).
56. S. Hadden, An integrable model for the dynamics of planetary mean-motion resonances. *Astron. J.* **158**, 238 (2019).
57. DC. Fabrycky *et al.*, Architecture of Kepler's multi-transiting systems. II. New investigations with twice as many candidates. *Astrophys. J.* **790**, 146 (2014).
58. L. M. Weiss *et al.*, The California-Kepler survey. V. Peas in a pod: Planets in a Kepler multi-planet system are similar in size and regularly spaced. *Astron. J.* **155**, 48 (2018).
59. B. V. Chirikov, A universal instability of many-dimensional oscillator systems. *Phys. Rep.* **52**, 263–379 (1979).
60. J. Bergstra, D. Yamins, D. D. Cox, Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *Proceedings of the 30th International Conference on International Conference on Machine Learning* (Association for Computing Machinery, 2013), **Vol. 28**, pp. I-115–I-123.
61. A. C. Petit, J. Laskar, G. Boué, Hill stability in the AMD framework. *Astron. Astrophys.* **617**, A93 (2018).
62. K. Batygin, K. M. Deck, M. J. Holman, Dynamical evolution of multi-resonant systems: The case of GJ 876. *Astron. J.* **149**, 167 (2015).
63. J. Chen, D. Kipping, Probabilistic forecasting of the masses and radii of other worlds. *Astrophys. J.* **834**, 17 (2016).
64. S. Chatterjee, E. B. Ford, S. Matsumura, F. A. Rasio, Dynamical outcomes of planet-planet scattering. *Astrophys. J.* **686**, 580–602 (2008).
65. M. Jurić, S. Tremaine, Dynamical origin of extrasolar planet eccentricity distribution. *Astrophys. J.* **686**, 603–620 (2008).
66. R. P. Butler *et al.*, Catalog of nearby exoplanets. *Astrophys. J.* **646**, 505 (2006).
67. V. Van Eylen, S. Albrecht, Eccentricity from transit photometry: Small planets in Kepler multi-planet systems have low eccentricities. *Astrophys. J.* **808**, 126 (2015).
68. J. W. Xie *et al.*, Exoplanet orbital eccentricities derived from Lamost–Kepler analysis. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11431–11435 (2016).
69. C. J. Burke *et al.*, Terrestrial planet occurrence rates for the Kepler GK dwarf sample. *Astrophys. J.* **809**, 8 (2015).
70. M. J. Holman, N. W. Murray, The use of transit timing to detect terrestrial-mass extrasolar planets. *Science* **307**, 1288–1291 (2005).
71. S. Hadden, T. Barclay, M. J. Payne, M. J. Holman, Prospects for TTV detection and dynamical constraints with TESS. *Astrophys. J.* **158**, 146 (2019).
72. Y. Lithwick, J. Xie, Y. Wu, Extracting planet mass and eccentricity from TTV data. *Astrophys. J.* **761**, 122 (2012).
73. D. Jontof-Hutter *et al.*, Secure mass measurements from transit timing: 10 Kepler exoplanets between 3 and 8 MEarth with diverse densities and incident fluxes. *Astrophys. J.* **820**, 39 (2016).
74. K. Volk, R. Malhotra, Dynamical instabilities in systems of multiple short-period planets are likely driven by secular chaos: A case study of Kepler-102. arXiv:2003.05040 (11 March 2020).
75. R. Barnes, T. Quinn, The (in) stability of planetary systems. *Astrophys. J.* **611**, 494 (2004).
76. S. Tremaine, The statistical mechanics of planet orbits. *Astrophys. J.* **807**, 157 (2015).
77. A. Izidoro *et al.*, Breaking the chains: Hot super-Earth systems from migration and disruption of compact resonant chains. *Mon. Not. Roy. Astron. Soc.* **470**, 1750–1770 (2017).
78. N. Murray, M. Holman, Diffusive chaos in the outer asteroid belt. *Astron. J.* **114**, 1246–1259 (1997).
79. D. Tamayo *et al.*, Predicting the long-term stability of compact multi-planet systems. Zenodo. https://zenodo.org/record/3723292#.XwcwoShKgdU. Deposited 21 March 2020.
80. B. Gladman, Dynamics of systems of two close planets. *Icarus* **106**, 247 (1993).
81. S. Hadden, Y. Lithwick, Numerical and analytical modeling of transit timing variations. *Astrophys. J.* **828**, 44 (2016).
82. H. Rein, S. F. Liu, Rebound: An open-source multi-purpose N-body code for collisional dynamics. *Astron. Astrophys.* **537**, A128 (2012).
83. H. Rein, D. Tamayo, A new paradigm for reproducing and analysing N-body simulations of planetary systems. *Mon. Not. Roy. Astron. Soc.* **467**, 2377 (2017).