The Astrophysical Journal, 944:190 (7pp), 2023 February 20 https://doi.org/10.3847/1538-4357/acb6ff

Simple Physics and Integrators Accurately Reproduce Mercury Instability Statistics

Dorian S. Abbot[1], David M. Hernandez[2,3], Sam Hadden[4], Robert J. Webber[5], Georgios P. Afentakis[6], and Jonathan Weare[7]
[1] Department of the Geophysical Sciences, The University of Chicago, Chicago, IL 60637, USA; abbot@uchicago.edu
[2] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[3] Department of Astronomy, Yale University, New Haven, CT 06511, USA
[4] Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, ON M5S 3H8, Canada
[5] Department of Computing & Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA
[6] Department of Physics, The University of Chicago, Chicago, IL 60637, USA
[7] Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA
Received 2022 December 12; revised 2023 January 26; accepted 2023 January 27; published 2023 February 27

Abstract

The long-term stability of the solar system is an issue of significant scientific and philosophical interest. The mechanism leading to instability is Mercury's eccentricity being pumped up so high that Mercury either collides with Venus or is scattered into the Sun. Previously, only three five-billion-year *N*-body ensembles of the solar system with thousands of simulations have been run to assess long-term stability. We generate two additional ensembles, each with 2750 members, and make them publicly available at https://archive.org/details/@dorianabbot. We find that accurate Mercury instability statistics can be obtained by (1) including only the Sun and the eight planets, (2) using a simple Wisdom–Holman scheme without correctors, (3) using a basic representation of general relativity, and (4) using a time step of 3.16 days. By combining our solar system ensembles with previous ensembles, we form a 9601-member ensemble of ensembles. In this ensemble of ensembles, the logarithm of the frequency of a Mercury instability event increases linearly with time between 1.3 and 5 Gyr, suggesting that a single mechanism is responsible for Mercury instabilities in this time range and that this mechanism becomes more active as time progresses. Our work provides a robust estimate of Mercury instability statistics over the next five billion years, outlines methodologies that may be useful for exoplanet system investigations, and provides two large ensembles of publicly available solar system integrations that can serve as test beds for theoretical ideas as well as training sets for artificial intelligence schemes.

Unified Astronomy Thesaurus concepts: Solar system (1528); Mercury (planet) (1024); Planetary dynamics (2173)

1. Introduction

The long-term stability of the solar system is important for our fundamental understanding of solar system dynamics, for contextualizing the solar system in comparison to exoplanet systems, and for evaluating the validity of the Copernican Principle. The most likely identified route to solar system instability is Mercury's eccentricity being pumped up to a high enough value that it has a close encounter with Venus (Laskar 1994; Laskar & Gastineau 2009), either colliding with Venus or being scattered into the Sun (Zeebe 2015). Mercury instability events are ultimately due to resonances of the secular system's modes (Lithwick & Wu 2011, 2014; Boué et al. 2012; Batygin et al. 2015; Mogavero & Laskar 2021, 2022).

The probability that Mercury's orbit becomes unstable has been estimated by three previous ensembles of *N*-body simulations (first by Laskar & Gastineau 2009 and subsequently by Zeebe 2015; Abbot et al. 2021) with thousands of members integrated for five billion years using symplectic integrators (Hairer et al. 2006). Brown & Rein (2020) ran a 96-member ensemble that resulted in no Mercury instability events, Brown & Rein (2022) ran a multi-thousand-member ensemble including the effects of perturbations by passing stars, and there have been a few other very small ensembles (e.g., Ito & Tanikawa 2002; Mikkola & Lehto 2022). Each of the large solar system ensembles was run with slightly different physical assumptions, numerical schemes, and time steps (Table 1). We will show in Section 3 that the Mercury instability statistics of the Laskar & Gastineau (2009) and Zeebe (2015) ensembles do not differ statistically, and that the Mercury instability statistic of the Abbot et al. (2021) ensemble is significantly larger in the range of 4–4.75 billion years.

Hernandez et al. (2022) showed that all existing ensembles are subject to numerical chaos due to time stepping, and this artificial instability can be mitigated by using smaller time steps. They postulated that the resulting artificial diffusion could affect phase-space statistics such as the Mercury instability probability. To avoid numerical chaos, they suggested a conservative time step that is $\frac{1}{16}$ of the effective period at pericenter of the innermost planet, following Wisdom (2015):

$$dt = \frac{\pi}{8}\sqrt{\frac{(1-e)^3}{1+e}\frac{a^3}{GM}}, \tag{1}$$

where e is eccentricity, a is semimajor axis, M is the self-gravitating mass, and G is the gravitational constant. Close encounters between Mercury and Venus occur when Mercury's eccentricity is about 0.85, which corresponds to a time step of 0.23 days according to Equation (1).

In this paper, we present two new large ensembles (2750 members) of the solar system for the next five billion years. Except for the integrator (see Section 2), both use a similar numerical setup as Brown & Rein (2020, 2022) and Abbot et al. (2021), but the first (`Fix dt`) has a fixed time step of $\sqrt{10} \approx 3.16$ days, and the second (`Var dt`) uses a time step

Table 1
Comparison of the Physical Assumptions and Numerical Schemes Used in the Three Previous Large Ensembles of Solar System Integrations for Five Billion Years

Ensemble	Bodies Included	Extra	General Relativity	Integrator	Time Step
Laskar & Gastineau (2009)	Sun, eight planets, Pluto, Earth's moon	Earth–moon tidal dissipation, Solar quadrupole moment	Saha & Tremaine (1994)	Laskar & Robutel (2001)	9.13 days
Zeebe (2015)	Sun, eight planets, Pluto		Saha & Tremaine (1994)	Rauch & Hamilton (2002)	4 days
Abbot et al. (2021)	Sun, eight planets		Nobili & Roxburgh (1986)	Wisdom et al. (1996)	8.06 days

Notes. Laskar & Gastineau (2009) reduce their time step for $e > 0.4$ but do not say to what value. Additionally, because of the scheme they use, their time step is effectively 0.577 times the stated value (Hernandez et al. 2022). Zeebe (2015) reduces his time step to 1 day for $0.55 < e < 0.70$ and 0.25 days for $0.70 < e < 0.80$.

that shrinks as Mercury's eccentricity increases, so as never to violate the Wisdom (2015) and Hernandez et al. (2022) criterion (Equation (1)). The Mercury instability statistics of both `Fix dt` and `Var dt` ensembles are consistent with those of Laskar & Gastineau (2009) and Zeebe (2015) (Section 3), suggesting that the Mercury instability statistics are robust to fairly large differences in physical assumptions and numerical schemes (Table 1). In Section 4, we perform convergence tests related to the Mercury instability problem and find that a time step of ≈9 days is sufficient to ensure convergence for these tests. We discuss our results in Section 5 and conclude in Section 6.

2. Model

We perform all simulations using the REBOUND *N*-body code's (Rein & Liu 2012) `WHFAST` integration scheme (Rein & Tamayo 2015), which is a Wisdom–Holman scheme (WH; Wisdom & Holman 1991). We do not use REBOUND's `WHCKL` scheme, as Brown & Rein (2020, 2022), Abbot et al. (2021) did, based on results from Hernandez et al. (2020, 2022), suggesting there is no clear advantage to a higher-order scheme for this problem. We emphasize that we have not used symplectic correctors (Wisdom et al. 1996). An integration improved by correctors indicates numerical instability is not present (Wisdom 2015; Hernandez et al. 2022), but correctors do not appreciably affect the existence of numerical instability. As in Abbot et al. (2021), we define a Mercury instability event as occurring when Mercury passes within 0.01 au of Venus, and we stop the simulations at that point. We do not attempt to resolve the close encounter of Mercury and Venus or subsequent behavior.

Our integrations contain all eight solar system planets (except for the simulations with massless Mercuries in Section 4) and include an approximation of general relativity with a modified position-dependent potential (Nobili & Roxburgh 1986), which is implemented as the `gr_potential` scheme in REBOUNDx (Tamayo et al. 2020). We initialize the simulations with solar system conditions on 2018 February 10 from the NASA Horizons database using standard REBOUND functionality. We show that roundoff error remains small throughout our five-billion-year integrations in Appendix A.

We produce two ensembles of simulations (`Var dt` and `Fix dt`) that both contain 2750 members. In `Var dt`, we use the heartbeat functionality in REBOUND to halve the time step if the Wisdom (2015) and Hernandez et al. (2022) criterion (Equation (1)) is about to be violated. Specifically, Mercury's heliocentric distance, r_M, is monitored, and the time step, dt, is halved whenever $r_M < a_M(1 - e_{\rm crit}(dt))$ where $a_M \approx 0.387$ au is Mercury's semimajor axis, which we approximate as fixed, and $e_{\rm crit}(dt)$ is the critical eccentricity satisfying Equation (1) for the current time step. We generate the `Var dt` ensemble by adding Gaussian perturbations to Mercury's x-position with a 1 cm scale. Because the Gaussian perturbations are so small and because of limited numerical precision, 510 of the simulations are identical to another simulation. Because the repeated simulations are unbiased with respect to Mercury instability events, we include them in our statistics. In the `Fix dt` ensemble, we use a fixed time step of $\sqrt{10} \approx 3.16$ days. We initialize the `Fix dt` ensemble on a uniform grid of perturbations to Mercury's x-position, each separated by 10 cm. This avoids the duplicated simulation problem encountered with our initialization of the `Var dt` ensemble.

Our `Var dt` and `Fix dt` ensembles are publicly available at https://archive.org/details/@dorianabbot. We save the model state every 10,000 yr throughout the simulations using the REBOUND simulation archive feature (Rein & Tamayo 2017), resulting in a total of 5 TB of data.

3. Solar System Ensembles

Figure 1 contains Mercury instability statistics as a function of time for the three previous large *N*-body solar system ensembles (Laskar & Gastineau 2009; Zeebe 2015; Abbot et al. 2021) and for our two new ones, one with a variable time step (`Var dt`) and one with a fixed time step (`Fix dt`). The author of Zeebe (2015) shared his data with us, and we obtained the instability times from Laskar & Gastineau (2009) using a plot digitizer. We estimate 1σ error bars for all ensembles, using the fact that the observed instability events are binomial random variables (Abbot et al. 2021).

Despite differences in physical and numerical assumptions, the ensembles produce broadly similar results. All ensembles show a roughly exponential increase in the number of events after three billion years, from a probability of $\sim 10^{-3}$ at three billion years to $\sim 10^{-2}$ at five billion years. Before three billion years, the ensembles have two or fewer Mercury instability events, which is too few to produce a discernible pattern. Our `Var dt` and `Fix dt` ensembles produce Mercury instability events before three billion years, like those of Laskar & Gastineau (2009) and Zeebe (2015). Our `Fix dt` ensemble produces the earliest Mercury instability event of any of the ensembles (Figure 2). The main reason the first Mercury instability event from our ensemble from Abbot et al. (2021) occurs at 2.95 billion years is likely that the ensemble only had 1008 members. We were able to produce an ensemble with earlier Mercury instability events in Abbot et al. (2021) using a rare-event simulation algorithm.

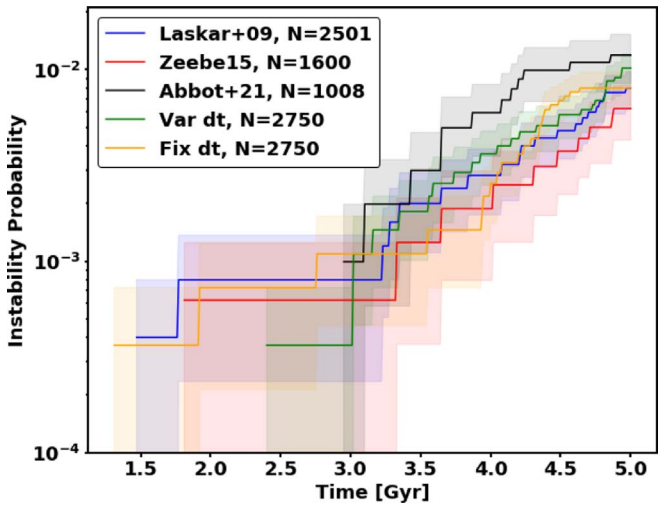

Figure 1. Mercury instability probability as a function of time for the Laskar & Gastineau (2009) ensemble (2501 members; blue), the Zeebe (2015) ensemble (1600 members; red), the Abbot et al. (2021) ensemble (1008 members, black), and the `Var dt` (2750 members, green) and `Fix dt` (2750 members, orange) ensembles from this paper. The shading represents 1σ error bars, using the fact that the observed instability events are binomial random variables.

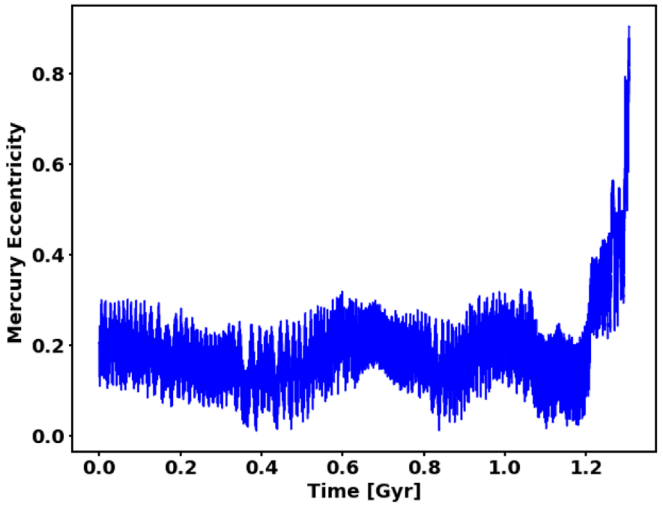

Figure 2. Mercury's eccentricity as a function of time for the simulation from the `Fix dt` ensemble in which Mercury's orbit becomes unstable the fastest.

It is tempting to try to interpret the differences among the ensembles. Before doing so, it is important to determine whether any of the differences are statistically significant. To do this, we need a statistical test to compare the rate of Mercury instability events in two populations (two different ensembles). Here, we use the statistical test that is recommended in D'agostino et al. (1988). Assume that the true proportion of Mercury instability events in population 1 is p_1 and the proportion in population 2 is p_2. The null hypothesis is that $p_1 = p_2$, or H_0: $p_1 = p_2$. The alternative hypothesis is $p_1 \neq p_2$, or H_a: $p_1 \neq p_2$. Let us take a random draw of n_1 members from population 1 and observe x_1 events for a sampled rate of $\hat{p}_1 = \frac{x_1}{n_1}$. Let us take a random draw of n_2 members from population 2 and observe x_2 events for a sampled rate of $\hat{p}_2 = \frac{x_2}{n_2}$. Now define the overall sample proportion as

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}. \quad (2)$$

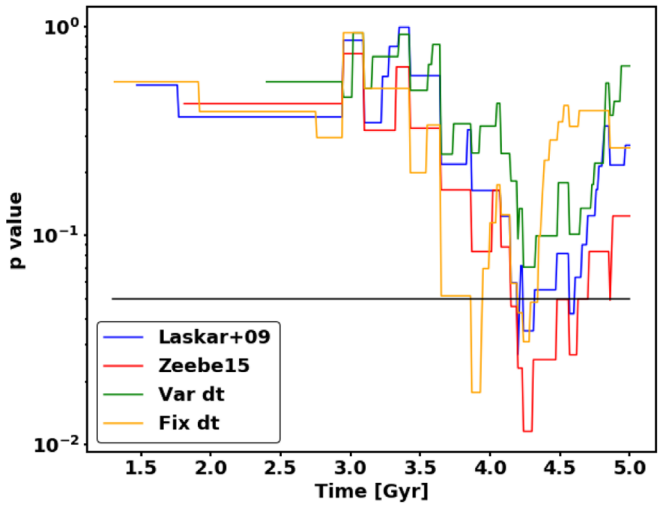

Figure 3. p-value testing the null hypothesis that the Mercury instability event rate of the plotted ensemble is equal to the rate of the Abbot et al. (2021) ensemble, using the test described in Equations (2)–(3). The black line represents $p = 0.05$. The ensembles tested are the Laskar & Gastineau (2009) ensemble (blue), the Zeebe (2015) ensemble (red), the `Var dt` ensemble (green), and `Fix dt` ensemble (orange).

We form the z statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}}, \quad (3)$$

where our estimate of the standard error is

$$\hat{\sigma} = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}. \quad (4)$$

Because we have no reason to believe either population is larger a priori, we need to do a two-tailed test using this z statistic.

When we perform this test on each combination of the ensembles of Laskar & Gastineau (2009), Zeebe (2015), `Var dt`, and `Fix dt`, we find that they are statistically consistent over the entire five billion year period tested. This is not surprising, given the overlap of the 1σ error bars in Figure 1. Figure 3 shows the results of the statistical test comparing the Abbot et al. (2021) ensemble to the other four ensembles. The Abbot et al. (2021) ensemble is statistically different from all of them ($p = 0.05$), except `Var dt` at some point between 3.75 and 4.75 billion years. This is consistent with the claim that the time step in the Abbot et al. (2021) ensemble is too large (Hernandez et al. 2022).

Because we cannot distinguish among the Laskar & Gastineau (2009), Zeebe (2015), `Var dt`, and `Fix dt` ensembles statistically, we can combine them to form one large ensemble of ensembles with 9601 members (Figure 4). The logarithm of the probability that a Mercury instability event is approximately linear in time between 1.3 and 5 Gyr and can be fit to the line

$$\log_{10} P = A + BT, \quad (5)$$

where $A = -0.413 \pm 0.013$ and $B = 0.469 \pm 0.004$. This indicates that the probability of a Mercury instability event grows exponentially by almost an order of magnitude for every two billion years of time elapsed. If Equation (5) holds true for times earlier than 1.3 Gyr, it suggests that the probability of a

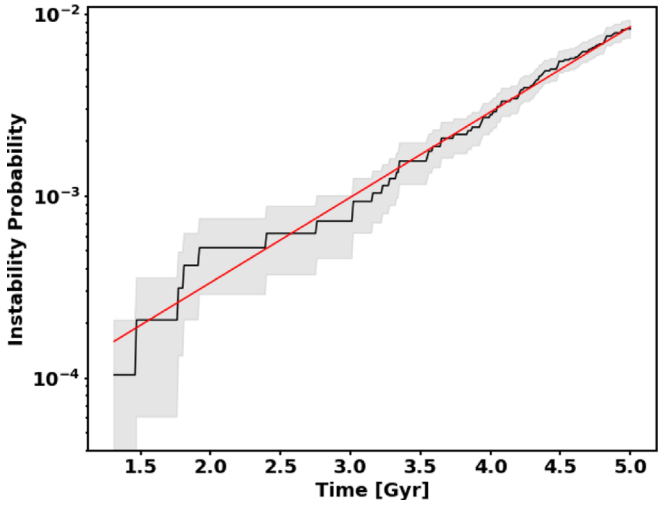

Figure 4. Mercury instability probability as a function of time for the combined ensemble of ensembles including Laskar & Gastineau (2009), Zeebe (2015), `Var dt`, and `Fix dt` (9601 members). The shading represents 1σ error bars, using the fact that the observed instability events are binomial random variables. The red line shows the line of best fit for the logarithm of the probability of a Mercury instability event as a function of time in billions of years (Equation (5)).

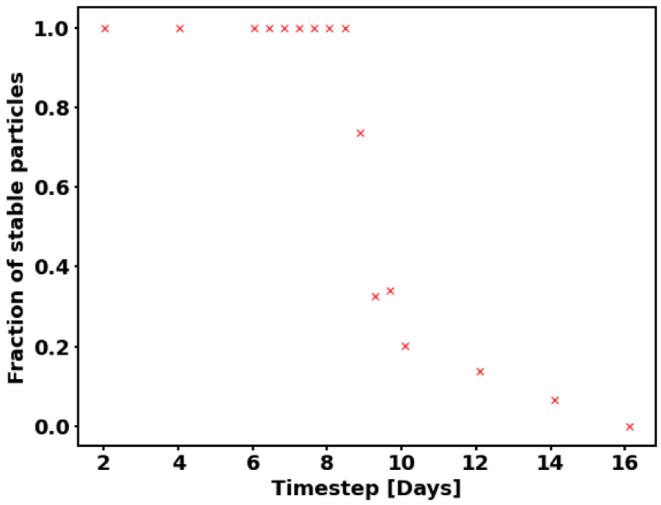

Figure 5. Fraction of 1000 massless ghost Mercuries that survive for one million years in a solar system integration as a function of time step (red cross symbols).

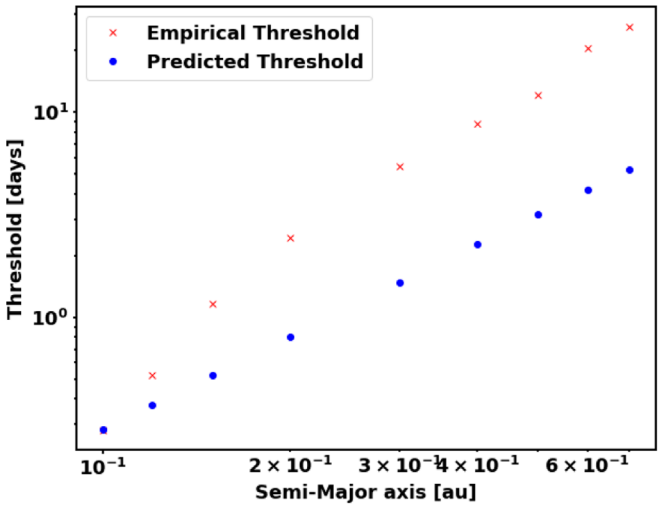

Figure 6. Our empirical time step threshold for convergence for the simulations with 1000 massless ghost Mercuries at various fixed semimajor axes (red cross symbols), as well as the predicted time step threshold assuming an eccentricity of 0.4 (Equation (1); blue circles).

Mercury instability event is about 6.6×10^{-5} in the next 0.5 billion years and about 1.1×10^{-4} in the next one billion years.

4. Convergence Tests

In this section, we perform two convergence tests relevant to Mercury instability events using the Wisdom–Holman scheme as implemented in the REBOUND *N*-body code. In the first test, we run simulations of the solar system with one thousand massless versions of Mercury and determine the time step necessary to get numerically stable results. In the second test, we do the same thing using one thousand massless versions of Mercury at different semimajor axes. In both tests, we find that a time step of about 9 days is small enough to get converged ensemble statistics. This is much larger than the 0.23 day criterion predicted by Wisdom (2015) and Hernandez et al. (2022) (Equation (1)).

In our first convergence test, we run simulations of the solar system with 1000 massless ghost Mercuries that are exact copies of Mercury except that they have zero mass and are initialized with true anomalies linearly spaced from 0 to 2π. We perform one-million-year simulations and study the fraction of ghost Mercuries whose eccentricity never exceeds 0.8. This fraction is 1 for a sufficiently small time step. We find that the model has converged at a time step of 9.1 days (Figure 5). We extended some of our simulations for 10 million years and found similar results.

In our second convergence test, we run simulations of a modified solar system with the Sun and all of the planets except Mercury. We then add 1000 massless ghost Mercuries at a value of the semimajor axis (a) that we vary among the different ensembles. All ghost Mercuries start with an eccentricity of 0.2 and are initialized with true anomalies linearly spaced from 0 to 2π. For each value of a, we perform one-million-year simulations and study the fraction of ghost Mercuries whose eccentricity never exceeds 0.8. For $a < 0.6$, this fraction is 1 for a sufficiently small time step. For $a = 0.6$ and $a = 0.7$, we remove Venus and again this fraction is 1 for a sufficiently small time step. We consider an ensemble converged at the largest time step for which the fraction of stable ghost Mercuries is 1.

In Figure 6, we plot the empirical time step threshold for convergence as a function of the semimajor axis of the ghost Mercuries. For $a = 0.1$, we find that Equation (1) is a good predictor of empirical time step threshold if we use an eccentricity of 0.4, under the assumption that most ghost Mercuries that reach this eccentricity will become unstable quickly. As a increases, the empirical time step threshold grows relative to Equation (1), such that it is 3–4 times larger than the predicted time step threshold for $a \gtrsim 0.3$. For these values of a, we find a slope in the log–log plot of empirical time step threshold versus semimajor axis with a 95% confidence interval of 1.5–2.3, which is consistent with the prediction of $a^{1.5}$ in Equation (1), although possibly higher. Finally, for $a = 0.4$, which is the closest semimajor axis we tested to Mercury's true value, we find an empirical time step threshold of 8.8 days.

5. Discussion

It might seem surprising that our `Fix dt` and `Var dt` ensembles produce Mercury instability statistics indistinguishable from those of Laskar & Gastineau (2009) and Zeebe (2015) (Figure 1), given the seemingly large differences in physical assumptions and numerical schemes (Table 1). This suggests that physical factors such as Earth's Moon, Pluto, Earth–Moon tidal dissipation, and the Solar quadrupole moment do not significantly affect Mercury instability statistics. Moreover, although including general relativity has an enormous effect on Mercury instability statistics (Laskar & Gastineau 2009), the detailed specification of general relativity does not have a large impact. Finally, it is reasonable that the order of the Wisdom–Holman scheme does not impact Mercury instability statistics, given that Hernandez et al. (2022) showed that higher-order Wisdom–Holman methods do not impact the onset of numerical chaos.

This work raises the question of why the Wisdom (2015) and Hernandez et al. (2022) criterion (Equation (1)) predicts a time step more than 1 order of magnitude smaller than what we find is necessary for convergence of ensemble Mercury instability statistics. A major part of the explanation is likely that it is not necessary to correctly resolve Mercury's orbit all the way until a collision or scattering with Venus in order to correctly infer statistics in an ensemble. Most ensemble members in which Mercury's eccentricity exceeds 0.4 for any significant period of time experience a Mercury instability event eventually (Laskar & Gastineau 2009; Zeebe 2015; Abbot et al. 2021). Equation (1) predicts a time step of 2.2 days for an eccentricity of 0.4, which is almost 10 times larger than the 0.23 day it predicts for an eccentricity of 0.85. A time step of 2.2 days is similar to the time step of 3.16 days that we used in the `Fix dt` ensemble that produced results statistically identical to those of the `Var dt` ensemble. It is interesting to note that we found a criterion of about 9 days for convergence in Section 4. This suggests that it might be possible to get converged results with a somewhat larger fixed time step. We should also note, however, that failure in the convergence tests we performed in Section 4 was defined by numerical noise dominating the system on a short timescale. It would not be surprising if a more stringent constraint on the time step is necessary to reproduce more nuanced physical variables in 5 Gyr integrations. Finally, the fact that Mercury instability statistics converge at a time step more than 1 order of magnitude larger than the conservative Wisdom (2015) and Hernandez et al. (2022) criterion (Equation (1)) does not imply that the same is true for all other statistics of other functions of phase space.

If we consider a single one of the Laskar & Gastineau (2009), Zeebe (2015), `Var dt`, and `Fix dt` ensembles, it might appear that Mercury instability events have different characteristics before and after 3 Gyr, potentially pointing to different instability mechanisms at different timescales. However, we find that a single line (Equation (5)) fits well the logarithm of the probability of Mercury instability events as a function of time for the ensemble of these ensembles between 1.3 and 5 Gyr. This suggests that a single mechanism is responsible for Mercury instability events between 1.3 and 5 Gyr, and that this mechanism becomes more pronounced with time.

Using the line of best fit (Equation (5)), we can estimate the probability of a Mercury instability event occurring in the next two billion years to be 3.3×10^{-4}. In Abbot et al. (2021), we estimated the probability of a Mercury instability event occurring in the next two billion years using QDMC, a rare-event sampling algorithm. Our estimate here is consistent with our estimate in Abbot et al. (2021) when we used a target time for QDMC of 2.8 billion years. Our estimate here is about 1 order of magnitude higher than our estimate in Abbot et al. (2021) when we used a target time for QDMC of 2.4 billion years, but we only observed one event in the first two billion years in that ensemble, so we would expect high variance in that estimate.

The large data set of solar system futures we have generated could be useful for further data analysis and training artificial intelligence schemes. For example, the data set could be mined to identify system characteristics that predict Mercury instability events and the time horizon for meaningful predictions. Techniques such as this have already been used in ocean-atmosphere dynamics (Tantet et al. 2015; Chattopadhyay et al. 2020; Finkel et al. 2020, 2021, 2022; Wang et al. 2020; Jacques-Dumas et al. 2022; Miloshevich et al. 2022) and chemistry (Ma & Dinner 2005; Thiede et al. 2019).

6. Conclusions

The main conclusions of this paper are:

1. We produce two new publicly available 2750-member ensembles for the future evolution of the planets in the solar system for five billion years. When we combine these ensembles with previous ensembles, we find that the logarithm of the probability of Mercury instability events is linear in time between 1.3 and five billion years. This suggests a single mechanism that becomes more pronounced with time is responsible for Mercury instability events in this time frame.
2. Accurate Mercury instability statistics over the next five billion years can be obtained using a simple Wisdom–Holman scheme without correctors and a representation of general relativity, applied to the Sun and the eight planets.
3. A time step of 3.16 days is sufficient for numerical stability and converged ensemble Mercury instability statistics.
4. One of our ensembles produced a Mercury instability event 1.3 billion years in the future, which is the soonest ever observed with an *N*-body simulation.

We thank Hanno Rein for helpful discussions about numerical schemes. We thank Richard Zeebe for sharing simulation data with us. We thank an anonymous reviewer, Richard Zeebe, and Jacques Laskar for comments on a draft of this manuscript. This work was completed with resources provided by the University of Chicago Research Computing Center. This work was supported by the NASA Astrobiology Program grant No. 80NSSC18K0829 and benefited from participation in the NASA Nexus for Exoplanet Systems Science research coordination network.

Software: This research made use of the open-source projects Jupyter (Kluyver et al. 2016), iPython (Pérez & Granger 2007), and matplotlib (Hunter 2007).

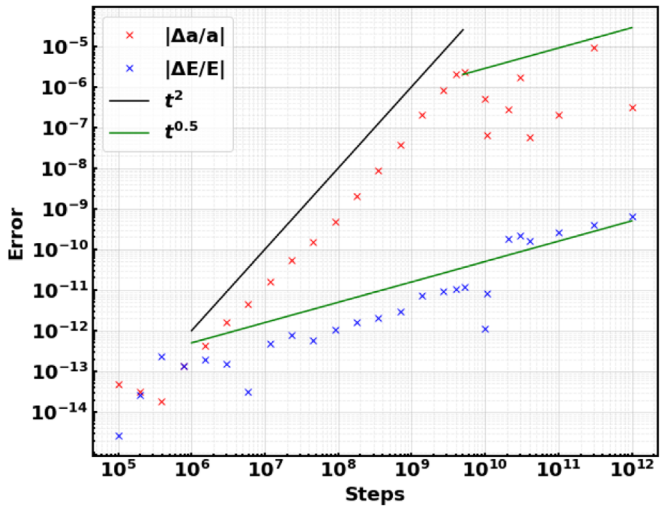

Figure 7. Accumulation of fractional error in total energy (E) and Mercury's semimajor axis (a) when a single realization of the solar system is run forward and backward for a given number of total time steps. The time step is 3.16 days and REBOUND is configured and initialized as described in Section 2. A 5 Gyr simulation corresponds to 5.8×10^{11} steps.

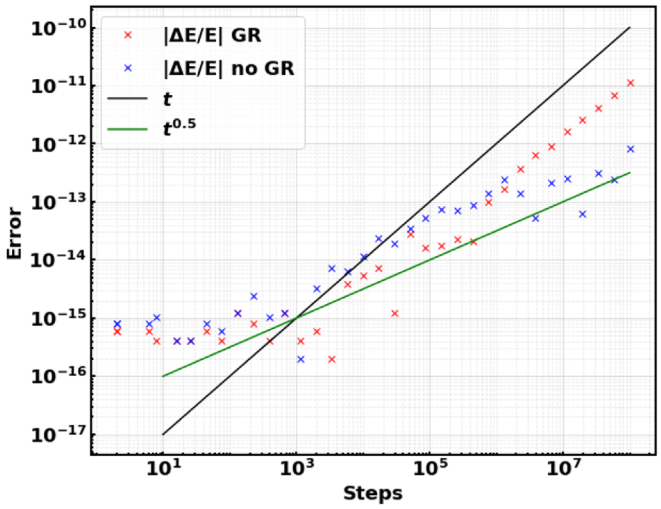

Figure 8. Accumulation of fractional error in energy including GR (red) and not including GR (blue) as a function of number of time steps for a REBOUND simulation including only the Sun and Mercury that is run forward only.

Appendix A
Roundoff Error Small in Our Simulations

Figure 7 shows the accumulated error in total energy and Mercury's semimajor axis when a single realization of the solar system is run forward and backward for a given number of total time steps using the REBOUND set up described in Section 2. In the absence of roundoff error, the errors would be 0. The maximum number of time steps we performed this test for is 10^{12}, which is larger than the 5.8×10^{11} steps necessary to run 5 Gyr with a 3.16 day time step. The energy error shows a simple $t^{0.5}$ scaling, indicating unbiased action-variable-like growth of roundoff error (Brouwer 1937). After initial unbiased behavior (see Figure 8), the error in Mercury's semimajor axis grows like t^2 until $10^{9.5}$ steps have passed, indicating biased angle-like roundoff error growth. After this, the error grows like $t^{0.5}$, indicating unbiased action-variable-like growth of roundoff error. The error both in the energy and Mercury's semimajor axis remain relatively small after 10^{12} steps, which suggests that roundoff error should not significantly affect our 5 Gyr integrations.

Appendix B
General Relativity and Biased Roundoff Error

The integrator WHFast has been shown to be unbiased for specific problems, numbers of steps, and phase-space variables (Rein & Tamayo 2015, 2017). However, it is possible that relativistic corrections affect this behavior. Here, we show how including a simple general relativistic potential can introduce bias in this integrator. For this test, we use Gaussian units, and consider a system composed only of the Sun and Mercury. The initial conditions are as follows:

```
sim.add(m=1.00000597682,
x=3.256101656448802E-03,
y=-1.951205394420489E-04,                z=-
1.478264728548705E-04,
vx=3.039963463108432E-06,
vy=6.030576499910942E-06,
vz=-7.992931269075703E-08) # Sun
```

```
sim.add(m=1.66051141e-7,                 x=-
1.927589645545195E-01,
y=2.588788361485397E-01,
z=3.900432597062033E-02,
vx=-2.811550184725887E-02,               vy=-
1.586532995282261E-02, vz=
1.282829413699522E-03) # Mercury.
```

The GR correction from Nobili & Roxburgh (1986) is used, and we calculate energy error over time with time step 0.0543988406 days. This is not a forward–backward integration as in Figure 7. We also compare against an integration without GR correction. The error grows as $\sqrt{N}$, where N is the number of time steps, when GR is neglected. However, a bias appears, and the growth of error scales with N, when GR is included. We tested a larger time step of 3.2 days and found that bias appears after a similar number of time steps as in Figure 8. The step number at which roundoff error becomes biased also agrees with the onset of bias in the semimajor axis of Mercury from our roundoff studies of the full solar system in Figure 7.

ORCID iDs

Dorian S. Abbot https://orcid.org/0000-0001-8335-6560
Robert J. Webber https://orcid.org/0000-0001-8286-6315

References

Abbot, D. S., Webber, R. J., Hadden, S., Seligman, D., & Weare, J. 2021, ApJ, 923, 236
Batygin, K., Morbidelli, A., & Holman, M. J. 2015, ApJ, 799, 120
Boué, G., Laskar, J., & Farago, F. 2012, A&A, 548, A43
Brouwer, D. 1937, AJ, 46, 149
Brown, G., & Rein, H. 2020, RNAAS, 4, 221
Brown, G., & Rein, H. 2022, MNRAS, 515, 5942
Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. 2020, JAMES, 12, e2019MS001958
D'agostino, R. B., Chase, W., & Belanger, A. 1988, Am. Stat., 42, 198
Finkel, J., Abbot, D. S., & Weare, J. 2020, JAtS, 77, 2327
Finkel, J., Webber, R. J., Gerber, E. P., Abbot, D. S., & Weare, J. 2021, MWRv, 149, 3647
Finkel, J., Webber, R. J., Gerber, E. P., Abbot, D. S., & Weare, J. 2022, JAtS, 80, 519

Hairer, E., Lubich, C., & Wanner, G. 2006, Geometrical Numerical Integration (2nd ed.; Berlin: Springer)
Hernandez, D. M., Hadden, S., & Makino, J. 2020, MNRAS, 493, 1913
Hernandez, D. M., Zeebe, R. E., & Hadden, S. 2022, MNRAS, 510, 4302
Hunter, J. D. 2007, CSE, 9, 90
Ito, T., & Tanikawa, K. 2002, MNRAS, 336, 483
Jacques-Dumas, V., van Westen, R. M., Bouchet, F., & Dijkstra, H. A. 2022, EGUsphere, egusphere–2022–1362
Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt (Amsterdam: IOS Press), 87
Laskar, J. 1994, A&A, 287, L9
Laskar, J., & Gastineau, M. 2009, Natur, 459, 817
Laskar, J., & Robutel, P. 2001, CeMDA, 80, 39
Lithwick, Y., & Wu, Y. 2011, ApJ, 739, 31
Lithwick, Y., & Wu, Y. 2014, PNAS, 111, 12610
Ma, A., & Dinner, A. R. 2005, JPCB, 109, 6769
Mikkola, S., & Lehto, H. J. 2022, CeMDA, 134, 1
Miloshevich, G., Cozian, B., Abry, P., Borgnat, P., & Bouchet, F. 2022, arXiv:2208.00971
Mogavero, F., & Laskar, J. 2021, A&A, 655, A1
Mogavero, F., & Laskar, J. 2022, A&A, 662, L3
Nobili, A. M., & Roxburgh, I. W. 1986, in IAU Symp. 114, Relativity in Celestial Mechanics and Astrometry, ed. J. Kovalevsky & V. A. Brumberg (Cambridge: Cambridge Univ. Press), 105
Pérez, F., & Granger, B. E. 2007, CSE, 9, 21
Rauch, K., & Hamilton, D. 2002, BAAS, 34, 938
Rein, H., & Liu, S.-F. 2012, A&A, 537, A128
Rein, H., & Tamayo, D. 2015, MNRAS, 452, 376
Rein, H., & Tamayo, D. 2017, MNRAS, 467, 2377
Saha, P., & Tremaine, S. 1994, AJ, 108, 1962
Tamayo, D., Rein, H., Shi, P., & Hernandez, D. M. 2020, MNRAS, 491, 2885
Tantet, A., van der Burgt, F. R., & Dijkstra, H. A. 2015, Chaos, 25, 036406
Thiede, E. H., Giannakis, D., Dinner, A. R., & Weare, J. 2019, JChPh, 150, 244111
Wang, X., Slawinska, J., & Giannakis, D. 2020, NatSR, 10, 2636
Wisdom, J. 2015, AJ, 150, 127
Wisdom, J., & Holman, M. 1991, AJ, 102, 1528
Wisdom, J., Holman, M., & Touma, J. 1996, Fields Inst. Commun., 10, 217
Zeebe, R. E. 2015, ApJ, 811, 9