REBOUNDx: A Library for Adding Conservative and Dissipative Forces To Otherwise Symplectic N-body Integrations

Daniel Tamayo$^1$, Hanno Rein$^{2,3}$, Pengshuai Shi$^3$, and David M. Hernandez$^{4,5,6}$

$^1$Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544
$^2$Department of Physical and Environmental Sciences, University of Toronto at Scarborough, Toronto, Ontario M1C 1A4, Canada
$^3$Department of Astronomy and Astrophysics, University of Toronto, Toronto, Ontario, M5S 3H4, Canada
$^4$Harvard-Smithsonian Center for Astrophysics, 60 Garden St., MS 51, Cambridge, MA 02138, USA
$^5$Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA
$^6$RIKEN Center for Computational Science, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, 650-0047 Hyogo, Japan

ABSTRACT

Symplectic methods, in particular the Wisdom-Holman map, have revolutionized our ability to model the long-term, conservative dynamics of planetary systems. However, many astrophysically important effects are dissipative. The consequences of incorporating such forces into otherwise symplectic schemes is not always clear. We show that moving to a general framework of non-commutative operators (dissipative or not) clarifies many of these questions, and that several important properties of symplectic schemes carry over to the general case. In particular, we show that explicit splitting schemes generically exploit symmetries in the applied external forces which often strongly suppress integration errors. Furthermore, we demonstrate that so-called ‘symplectic correctors’ (which reduce energy errors by orders of magnitude at fixed computational cost) apply equally well to weakly dissipative systems and can thus be more generally thought of as ‘weak splitting correctors.’ Finally, we show that previously advocated approaches of incorporating additional forces into symplectic methods work well for dissipative forces, but give qualitatively wrong answers for conservative but velocity-dependent forces like post-Newtonian corrections. We release REBOUNDx, an open-source C library for incorporating additional effects into REBOUND N-body integrations, together with a convenient PYTHON wrapper. All effects are machine-independent and we provide a binary format that interfaces with the SimulationArchive class in REBOUND to enable the sharing and reproducibility of results. Users can add effects from a list of pre-implemented astrophysical forces, or contribute new ones.

Key words: methods: numerical — gravitation — planets and satellites: dynamical evolution and stability

1 INTRODUCTION

The long-term dynamical evolution of planetary systems remains a rich challenge for analytical investigations. As a result, many of the advances in the last several decades have been driven by numerical studies thanks to faster computers and the development of improved algorithms for accurate and efficient numerical solutions, e.g., the chaotic evolution of Pluto (Sussman & Wisdom 1988), chaos in the inner solar system (Laskar 1989), and the marginal instability of Mercury over the age of the solar system Laskar & Gastineau (2009). In many cases, such chaos only manifests itself after millions or even billions of orbits. This imposes strong demands on the long-term conservation properties of numerical integration methods that can be used for such studies.

These concerns led to the development powerful symplectic integration techniques (e.g., Ruth 1983; Neri 1987; Forest & Ruth 1990; Yoshida 1990). Most notably, by exploiting the near-Keplerian planetary motions, the Wisdom-Holman map (Wisdom 1982; Kinoshita et al. 1990; Wisdom & Holman 1991) enabled the first direct N-body integrations of our solar system over Gyr timescales. By respecting the Hamiltonian structure of the problem, symplectic methods render the task of integration equivalent to a canonical transformation (e.g. Sanz-Serna 1992), which enforces the conservation of several invariants of the phase space flow. This avoids the secular energy error growth of many of the more general methods that are not restricted to Hamiltonian systems, which can lead to unphysical collisions in long-term simulations.
Such symplectic integrators split the problem into multiple pieces that are each evolved in sequence. Apart from the robust geometrical and numerical properties this provides, the modularity of splitting methods makes it possible to extend them for hierarchical problems (Fuji et al. 2007; Portegies Zwart & McMillan 2018) and to incorporate different physics in a single integration (Pelupessy & Portegies Zwart 2012; Portegies Zwart 2018). This has for example been vigorously pursued in the Astrophysics Multipurpose Software Environment (AMUSE) package (Portegies Zwart & McMillan 2018).

Some effects, such as bodies’ higher gravitational moments, can be expressed as simple position-dependent potentials and can be trivially incorporated into symplectic schemes. However, many crucial perturbations such as post-Newtonian corrections, tides, or radiation forces are velocity-dependent and even dissipative. Malhotra (1994) and Cordeiro et al. (1996) have proposed generalizations of symplectic schemes for incorporating weak velocity-dependent forces, but their consequences in long-term integrations are not well understood. While some authors perform convergence tests to check the validity of their adopted stepsize using variants of the above methods (e.g. Zhang & Hamilton 2007), the numerical robustness of studies where this is not done is unclear.

A generalization of symplectic integration theory would therefore be valuable to understand the source of numerical error in the above schemes, and to provide simple estimates for its magnitude to guide the selection of timestep for a given problem.

A promising direction has been offered by Galley (2013) and Galley et al. (2014), who have developed a generalized theory for classical mechanics that can incorporate dissipative processes in the action governing the dynamics. One can use this formalism to construct variational integrators that generalize symplectic methods to accurately track the changing momentum and energy of the system on long timescales (Tsang et al. 2015). Potential disadvantages of this elegant framework are that the equations of motion are necessarily implicit (Tsang et al. 2015), and that adding new forces becomes somewhat more complicated than with traditional methods, since one needs to derive a ‘nonconservative potential’ using the formalism in Galley et al. (2014).

In this paper, we explore how the established theory of symplectic integration can instead be generalized in a framework of non-commutative operators, which can be applied to both conservative and dissipative systems. Such generalized ‘splitting methods’ have been suggested in the field of fluid mechanics as far back as Strang (1968), and more recently for celestial mechanics by Mikkola (1998)1. We show that several strong results traditionally discussed in the symplectic integration literature carry over to more general and possibly dissipative splitting methods.

Readers interested in quickly using REBOUNDx to incorporate astrophysical effects into N-body integrations could begin at Sec. 6 for an overview of the implementation. For a more gradual introduction, we introduce in Sec. 2 the notation, and review symplectic integration and the Wisdom-Holman map in an operator-centred formalism. We then show in Sec. 3 how this framework can explain the energy error behaviour of such methods under weak perturbations, both conservative and dissipative. In Sec. 4 we review how the WH map can be extended to higher order without additional force evaluations, and show that such ‘symplectic correctors’ can also be applied to dissipative systems and are thus more general than typically considered. In Sec. 5, we consider the general case where analytic solutions to individual steps in a split scheme are not known, and how truncation errors from numerically integrating across these individual steps interact with the overall splitting scheme errors. In particular, we show that for conservative, velocity-dependent forces like post-Newtonian corrections, the methods of Cordeiro et al. (1996) and Malhotra (1994) give qualitatively wrong answers, and we show how to correct them. We summarize our results and conclude in Sec. 7.

2 BACKGROUND

We begin by reviewing the basic theory behind symplectic integration. Several excellent reviews are available (e.g. Kinosita et al. 1990; Sanz-Serna 1992), but we choose to present it in an operator-centred framework that will generalize later to dissipative systems. We try to keep the introduction at the level required for the discussion in the main text. For a more careful introduction, and to correct some sign errors in the literature, see Appendix A.

2.1 Splitting Schemes

We consider a system of N particles with positions ri and velocities vi. We define a differential operator ˘P, which acts on the current state of the system z = (r, v) to yield ˘z, i.e., the particular set of differential equations we are trying to solve2:

\[ ˘P \hat{z} : \quad \hat{r}_i = v_i, \quad \hat{v}_i = a_i(r), \quad (1) \]

where ai is the ith particle’s acceleration vector, which depends on the positions r of all the bodies.

Next, we introduce a further level of abstraction by defining a solution to Eq. 1 through a corresponding operator ˘P(h). This ideal integrator exactly advances the state z by a timestep h according to the set of differential equations ˘z = ˘Pz.

In general, there is no closed form solution ˘P (e.g., the three-body problem), but one could split the differential equations into two pieces that could each be solved trivially in isolation,

\[ \hat{A} \hat{z} : \quad \hat{r}_i = v_i, \quad \hat{v}_i = 0, \]
\[ \hat{B} \hat{z} : \quad \hat{r}_i = 0, \quad \hat{v}_i = a_i(r). \quad (2) \]

In particular, the solution ˘A(h) keeps the velocities constant and updates the positions by hv, while ˘B(h) keeps the positions constant and similarly updates the velocities using constant accelerations. This splitting works for any differential equations of the form in Eq. 1, and is thus widely applicable. In a Hamiltonian framework, in cases where the accelerations can be derived from a position-dependent potential, it corresponds to splitting the Hamiltonian into a kinetic piece containing all the momenta, and a potential piece containing all the positions. We therefore refer to this scheme as a kinetic-potential splitting, which was the focus of most early symplectic integrators (e.g., Ruth 1983; Neri 1987; Forest & Ruth 1990; Yoshida 1990). As we will see below, better splittings for specialized problems are possible.

The idea behind splitting methods is to alternate evolution under two or more operators that can be calculated exactly and efficiently. For example, we can define a first-order splitting scheme as

---

1 See also Hernandez & Bertschinger (2018) for error analysis of one-step methods from the perspective of their differential equations.

2 Throughout this paper we restrict ourselves to time-independent sets of differential equations.
\[ z(t + h) \approx S_{\mathcal{AB}}(h) z(t) = \mathcal{A}(h) \circ \mathcal{B}(h) z(t), \]  

where \( \mathcal{B} \) advances the state \( z(t) \) by \( h \) according to its half of the differential equations, and then \( \mathcal{A} \) advances the result by \( h \) with the remaining half. Unfortunately, doing two halves of the problem in sequence is not the same as doing the full problem all at once. Nevertheless, if we repeatedly apply this map to get from some \( z(t_i) \) to \( z(t_{i+1}) \), at least in the limit of a vanishing timestep, the split operations become densely interleaved and this approximation will approach the real solution.

For finite timesteps, we can also see simply that how well a split scheme will correspond to the real solution should depend on the degree to which the operators \( \mathcal{A}(h) \) and \( \mathcal{B}(h) \) commute with one another. We could always split each of the steps in Eq. 3 in half and write it as \( S_{\mathcal{AB}}(h) = \mathcal{A}(h/2) \circ \mathcal{B}(h/2) \circ \mathcal{B}(h/2) \circ \mathcal{A}(h/2) \). If the order in which these operators are applied did not matter, we could switch the middle two to yield \( \mathcal{A}(h/2) \circ \mathcal{B}(h/2) \circ \mathcal{A}(h/2) \circ \mathcal{B}(h/2) \). We could then continue this split-and-switch procedure ad infinitum until the operators were again densely interleaved and we would effectively be applying both at the same time, yielding the true solution. So we expect the errors in splitting schemes to vanish in the limits that either \( h \) or the commutator of \( \mathcal{A} \) and \( \mathcal{B} \) go to zero.

The formula that quantifies the above two statements is the Baker-Campbell-Hausdorff (BCH) identity, which is useful in many branches of physics involving non-commuting operators like quantum mechanics. To state it, we have to add a final layer of abstraction and generalize beyond just how the phase space variables \( z \) change under a particular subset of the differential equations (e.g., Eq. 2), to how an arbitrary function \( g(z) \) changes with time. In particular, for every subset of differential equations, e.g., \( \dot{z} = \mathcal{A} z \), we can define the corresponding Lie derivative \( \mathcal{L}_A \) that, when acting on \( g(z) \), yields \( dg/dt \) under that particular subset of differential equations. In the simple case of \( g(z) = z \), we get back \( \mathcal{L}_A z = z = \hat{A} z \).

The BCH formula says that when we apply Eq. 3, we are not solving the original set of differential equations \( \dot{z} = \mathcal{L}_A z + \mathcal{L}_B z \) (Eqs. 1 and 2), but rather a nearby set of differential equations

\[ \dot{z} = \mathcal{L}_{\mathcal{S}AB} z = \mathcal{L}_A z + \mathcal{L}_{\mathcal{S}AB} z, \]

where the additional error term is a power series in the timestep consisting of nested commutators \( [\mathcal{L}_A, \mathcal{L}_B] = \mathcal{L}_A \mathcal{L}_B - \mathcal{L}_B \mathcal{L}_A \) (e.g., Saha & Tremaine 1992),

\[ \mathcal{L}_{\mathcal{S}AB} = -\frac{h}{2} [\mathcal{L}_A, \mathcal{L}_B] + \frac{h^2}{12} [\mathcal{L}_A, [\mathcal{L}_A, \mathcal{L}_B]] - O(h^3). \]

In other words, if we integrate the set of differential equations in Eq. 4 numerically to high accuracy, we match the evolution generated by the simple splitting scheme in Eq. 3. We will call these differential equations that we are actually solving the ‘modified set of differential equations’ for the splitting scheme. Understanding the structure of this nearby problem gives insight into the numerical errors introduced by the method\(^3\).

As expected from the qualitative arguments above, the modified differential equations approach the true ones as \( h \) approaches zero. We also see that if (and only if) \( [\mathcal{L}_A, \mathcal{L}_B] = 0 \), then there is no error and \( \mathcal{A}(h) \) and \( \mathcal{B}(h) \) commute, i.e., \( \mathcal{A}(h) \circ \mathcal{B}(h) = \mathcal{B}(h) \circ \mathcal{A}(h) \).

By choosing appropriate steps, one can compose \( \mathcal{A} \) and \( \mathcal{B} \) into higher order schemes (e.g., Yoshida 1990). A widely used scheme is the time-symmetric ‘leapfrog’ method

\[ S_{\mathcal{AB}}(h) = \mathcal{A}(h/2) \circ \mathcal{B}(h) \circ \mathcal{A}(h/2). \]

Repeated application of the BCH formula shows that this particular composition cancels out error terms linear in \( h \), yielding a second-order integrator with

\[ \mathcal{L}_{\mathcal{S}AB} = \frac{h^2}{24} [2L_B + \mathcal{L}_A, [\mathcal{L}_B, \mathcal{L}_A]] + O(h^4). \]

Because the scheme is time-symmetric, errors only appear at even powers of \( h \).

2.2 Symplectic Schemes

The above analysis does not depend on us solving a Hamiltonian system, and is general to both conservative and dissipative systems. However, the insight of symplectic integration is that if each of the split operators, \( \mathcal{A}(h) \) and \( \mathcal{B}(h) \), are exact solutions to a Hamiltonian system, then each of them conserve various Hamiltonian invariants. This implies that any composition of them, e.g., \( \mathcal{A}(h) \circ \mathcal{B}(h) \), must also conserve those quantities. In particular, such schemes will conserve the Poincaré invariants, the linear and angular momentum, and the infinite differentiability class order of the governing differential equations (Hernandez 2019a), all of which are important for the accuracy of solutions (Hernandez 2019b). Having these strong conservation properties built into symplectic schemes gives them excellent long-term behaviour.

The main subtlety, however, is that each split operator, e.g., \( \mathcal{A}(h) \), is conserving its own Hamiltonian, e.g., \( H_A \). It is therefore not clear what Hamiltonian an overall composition scheme like the one in Eq. 3 is conserving. Then again, a Hamiltonian \( H_I \) yields the associated set of differential equations \( \dot{z} = \mathcal{L}_I z \) through Hamilton’s equations, so it should not be surprising that this answer can also be derived from the BCH formula.

One can show that one gets analogous formulae to Eqs. 5 and 7 (e.g., Yoshida 1993),

\[ H^{\text{err}}_{\mathcal{S}AB} = \frac{h^2}{2} [H_A, H_B] + \frac{h^4}{12} [H_A - H_B, [H_A, H_B]] + O(h^5), \]

and

\[ H^{\text{err}}_{\mathcal{S}AB} = \frac{h^2}{24} (2H_B + H_A, [H_B, H_A]) + O(h^4), \]

where curly brackets denote Poisson brackets. Thus, each of the error terms in the modified differential equations (e.g., Eq. 5) can be derived from a corresponding error Hamiltonian (e.g., Eq. 8) (e.g., Saha & Tremaine 1992). This is another way of seeing that the integration errors respect the symplectic geometry of the problem.

2.3 The Wisdom-Holman Map

While the energy errors in kinetic-potential splittings do not drift secularly with time, they remain large throughout the integration. Physically, this is because one is making drastic deviations from the true trajectory each timestep, i.e., force-free motion along \( \mathcal{A}(h) \) alternating with order-unity kicks in the particle velocities under \( \mathcal{B}(h) \).

In the case of weakly perturbed systems, like planets moving on nearly Keplerian orbits, it is much more powerful to instead split into an integrable dominant operator and the weak perturbation.
Such a scheme, today known as the Wisdom-Holman map, proved a major development for planetary integrations, enabling the first integrations of the solar system over Gyr timescales (Wisdom & Holman 1991).

While the development of the Wisdom-Holman map was not driven by this operator framework (see Wisdom 2018 for a historical perspective), it is instructive for our purposes to analyze it in this light. Instead of splitting the differential equations into components of comparable magnitudes like in Eq. 2, the Wisdom-Holman map splits them as (Wisdom & Holman 1991; Kinoshita et al. 1990),

$$L_h z : \begin{array}{ll} \dot{r}_i = v_i, & \dot{v}_i = \mathbf{a}_i^{\text{Kep}}(r) \\ \epsilon_p L_h z : \begin{array}{ll} \dot{r}_i = 0, & \dot{v}_i = \epsilon_p \mathbf{a}_i^{\text{int}}(r) \end{array} \end{array}$$

where the $\mathbf{a}_i^{\text{Kep}}$ are the two-body Keplerian accelerations, and the $\mathbf{a}_i^{\text{int}}$ are the remaining interaction accelerations$^4$. We have also introduced a factor of $\epsilon_p$ of order the characteristic planet-star mass ratio$^5$ to keep track of the fact that the interplanetary accelerations (and therefore the changes in the interaction steps) are much smaller than the dominant Keplerian accelerations due to the central star.

Like in the kinetic-potential split, the corresponding $I(h)$ is trivial, since it keeps the positions (and thus the accelerations) constant across the timestep. The exact solution for the velocities is then just

$$v_i(t + h) = v_i(t) + \epsilon_p h \mathbf{a}_i^{\text{int}}.$$

By contrast, the Kepler step $K(h)$ updates both the positions and velocities of each body along their respective unperturbed two-body orbits. Interestingly, more accurate and efficient schemes for solving the Kepler problem have continually been found for over three hundred years (e.g., Newton 1687; Machin 1737; Rambaut 1890; Plummer 1896; Brown 1931; Danby 1992; Mikkola & Innanen 1999; Rein & Tamayo 2015a; Wisdom & Hernandez 2015).

### 2.4 Energy Error Estimates

As an example, consider a case of two Earth-mass planets with semi-major axes of 1 and 2 AU around a solar-mass star ($\epsilon_p = 3 \times 10^{-6}$) with orbital eccentricities of 0.01. We specialize to second order schemes (Eq. 6), and compare the kinetic-potential splitting (Eq. 2) to the Wisdom-Holman splitting (Eq. 10). Figure 1 plots both methods’ fractional energy errors over ten thousand inner-planet orbits, using a timestep of 1% the innermost planet’s orbital period.

The Wisdom-Holman map conserves the energy many orders of magnitude better than the kinetic-potential splitting, and we can make simple order of magnitude estimates to understand why. This is clearest in a Hamiltonian framework for conservative systems (e.g. Wisdom & Holman 1991; Saha & Tremaine 1992), but we instead show how it arises in the above operator-centred picture that will later generalize to dissipative systems. While this is not a rigorous derivation, it is useful to be able to quickly estimate expected errors and their associated scalings. While we do not plot all possible dependencies, we have verified that the various scalings predicted below are correct.

A useful feature of the abstract Lie derivative formalism of Sec. 2 is that they describe not only how the phase space variables $z$ evolve, but also how functions like the energy $E(z)$, change with time. In particular, just like $E^{\text{K}} = L^{\text{K}}_q z$ yields the additional terms in the modified differential equations for the phase space variables using a particular split scheme $S$ (Sec. 2), $E^{\text{err}} = L^{\text{err}}_q E(z)$ yields the additional terms in the modified differential equations for the energy evolution.

To approximately estimate the error $\Delta E^{\text{err}}$ they cause, we can multiply the additional terms $E^{\text{err}}$ by the timescale over which these errors coherently add up. Toward this end, on the left of Fig. 1, the logarithmic time axis makes it possible to see that the errors accumulate coherently over several timesteps ($h = 0.01$), and that on longer timescales, they oscillate. We can understand this from Sec. 2. Because the BCH formula says that all the error terms in the modified differential equations are composed of the split operators that propagate the real dynamics (Eq. 7), the errors will inherit the timescales in the original problem. In particular, the timescale over which errors will accumulate coherently will be the shortest timescale in the problem. In this case of near-circular, non-resonant orbits, $T$ would be the synodic period on which planets kick each others’ orbits every conjunction. In our typical example where the orbits are not too close, this is approximately the orbital period of the innermost planet; for a very eccentric orbit, $T$ would be the much shorter timescale for pericentre passage (e.g., Rauch & Holman 1999; Rein & Tamayo 2015b), etc. We thus have

$$\frac{\Delta E^{\text{err}}}{E} \sim \frac{T E^{\text{err}}}{E} = \frac{T L^{\text{err}}_q E}{E},$$

where, for the second-order schemes used in Fig. 1, $L^{\text{err}}_q$ is given by Eq. 7.

Since $L_A$ and $L_B$ each yield time derivatives according to their own subset of the differential equations, we can make a rough estimate in Eq. 12 by replacing each instance of $L_A$ and $L_B$ in $L^{\text{err}}_q$ by their respective inverse characteristic timescale. In this case we

---

$^4$ These capture both the direct gravitational interactions between planets and their indirect effects on one another as they each pull on the central star. There are many ways to do this split between Keplerian and interaction accelerations (Hernandez & Dehnen 2017; Rein & Tamayo 2019). The splitting above into two operators corresponds to using Jacobi coordinates (Wisdom & Holman 1991).

$^5$ This assumes that the distances from the planets to the central star are comparable to their interplanetary separations. In general, $\epsilon_p$ would capture the ratio of the perturbation accelerations to those of the dominant Keplerian motion.
are referring specifically to the timescale on which each operator in isolation would change the particle states by order unity, i.e., \( \epsilon / 2 \).

For Kepler splittings, the Kepler Lie derivative \( \mathcal{L}_K \) induces order-unity changes on the orbital timescale \( T \). Perturbation operators like the interaction step in the WH map induce order-unity changes on timescales that are longer by roughly a factor of the perturbation strength \( \epsilon \) (e.g., for the interaction operator, \( \dot{v}/\dot{v} = (\dot{v}/\dot{v}_{K})/(\dot{v}/\dot{v}) = \epsilon_p T \), see Eq. 10).

If we further assume that the two operators \( \mathcal{L}_a \) and \( \mathcal{L}_b \) do not commute, i.e., \( [\mathcal{L}_a, \mathcal{L}_b] \sim 1 \), then plugging Eq. 7 into Eq. 12 yields a fractional energy error estimate of \( \sim \epsilon_p (h/T)^2 \). This gives the right scaling for a second-order method, as expected. Indeed, the same argument applied to any Kepler splitting says that if the perturbation accelerations are a factor of \( \epsilon \) smaller than the Keplerian ones, then the second-order splitting scheme given by Eq. 6 will yield energy errors \( \sim \epsilon h/T \).

By contrast, in the kinetic-potential splitting (Eq. 2), both \( \mathcal{L}_a \) and \( \mathcal{L}_b \) induce order-unity changes on the orbital timescale \( T \), so \( \mathcal{L}_a \sim \mathcal{L}_b \sim 1/T \). This yields a correspondingly larger error \( (h/T)^2 \).

More physically, the WH method is benefiting from the fact that the dominant motion is known and solved exactly. In the limit of \( \epsilon_p \rightarrow 0 \), the Wisdom-Holman map would yield an exact trajectory, while the kinetic-potential splitting would still be splitting the Kepler problem and the errors would remain large and unchanged.

These simple estimates (black arrow in Fig. 1) match well to integrations with the Wisdom-Holman map, while our estimate for the kinetic-potential splitting is depressed by approximately a factor of \( \epsilon \) from our simple estimate. This has to do with our simplification of the commutation relations, and has interesting broader implications that we discuss in Sec. 3.2.

In summary, while the Kepler step is more expensive to compute than the steps in a kinetic-potential splitting, the factor of \( \epsilon_p \) gain makes reaching a given level of accuracy significantly faster with the Wisdom-Holman map. Perhaps more importantly, in cases where inter-planetrary forces are very weak, the Wisdom-Holman method keeps the additional terms in the modified differential equations smaller than the interplanetary terms. By contrast, in the kinetic-potential splitting, these additional error terms can easily be larger than the inter-planetary forces, leading to spurious results.

These scalings suggest the cases where splitting methods will be most useful, i.e. when perturbations are weak \( \epsilon \ll 1 \) and when the shortest characteristic timescale \( T \) is not too short. For example, for a planet experiencing dynamical tides on an extremely eccentric orbit, the strength of the perturbation changes by orders of magnitude over an orbital period. In this case, the shortest timescale would be the characteristic timescale of pericentre passage, which can be much shorter than an orbital period. Because simple symplectic schemes require a fixed timestep, the small \( h \) required where the forces are strongest acts as a bottleneck for the remaining orbit that could otherwise be easily integrated. In such cases it is typically more efficient to use a more general scheme with adaptive timesteps like RAS or in some restricted cases it is possible to symplectically adapt inverse timesteps in inverse proportion to the strength of an external potential (Preto & Tremaine 1999; Mikkola & Tanikawa 1999; Petit et al. 2019).

3 ADDITIONAL FORCES

We now move beyond point-particle gravity to consider adding additional effects in N-body integrations. We will see that the above framework, by not specialising to Hamiltonian systems, can be used to straightforwardly understand the numerical behavior under both conservative and dissipative forces.

3.1 Conservative Forces

We begin by considering conservative forces that can be derived from position-dependent potentials. Hamiltonian perturbations that cannot be written strictly in terms of the particle positions are also important (e.g., post-Newtonian corrections), but we defer their discussion to Sec. 5.

We note that while any position-dependent potential can obviously be trivially incorporated into a kinetic-potential split, it can also be directly inserted into the interaction step in the Wisdom-Holman map as an additional position-dependent acceleration in Eq. 10. Then Eq. 11 remains the exact solution with the acceleration given by the sum of all the position-dependent accelerations in the problem.

As an example, we consider the same two-planet case as above, but now with additional perturbations from an oblate primary, with the planets orbiting in the primary’s equatorial plane. The first corrections to point source gravity in a multipole expansion of the primary’s potential is the quadrupole term, with a potential in the equatorial plane of

\[
V_{22}(r) = -\frac{1}{2} \left( \frac{R}{r} \right)^2 \frac{G M}{r} \equiv \epsilon_{22} \frac{G M}{a_0} \left( \frac{a_0}{r} \right)^3.
\]

The innermost body will be most affected, so we take \( a_0 \) to be its original, semimajor axis, \( G \) is the gravitational constant, \( M \) and \( R \) are the primary’s mass and radius, and \( J_2 \) is the standard dimensionless coefficient for the quadrupole field. We have also introduced a parameter \( \epsilon_{22} = \frac{1}{2} J_2 \left( \frac{a_0}{r} \right)^3 \) that captures the smallness of the effect relative to the dominant Keplerian potential \( \pm G M / a_0 \), since \( a_0 / r \) is approximately unity for nearly circular orbits. We take \( \epsilon_{22} \approx 10^{-3} \).

We again compare second-order schemes (Eq. 6) using kinetic-potential and Wisdom-Holman Kepler splittings in Fig. 2. Comparing with Fig. 1, we see that the kinetic-potential errors have not changed. This is because the planetary accelerations in \( \mathcal{L}_b \) for the kinetic-potential splitting (Eq. 2) are dominated by those due to the star. Adding a small acceleration due to the oblateness perturbation therefore does not noticeably change the dominant error terms in Eq. 7. By contrast, for the Kepler splitting, our chosen oblateness perturbations are much larger than the inter-planetary ones (\( \epsilon_{22} \gg \epsilon_p \)), so they change the error behavior. The same argument following Eq. 12 suggests an energy error \( \sim \epsilon_{22}^2 (h/T)^2 \).

However, both our error estimates for the kinetic-potential and Kepler splitting fall short by a factor of the eccentricity (both black arrows in Fig. 2 include this factor). This can be a significant suppression in the errors and is due to the geometrical properties of splitting methods, as we explore in the next section.

3.2 Splitting Methods Exploit Symmetries

Because for the remainder of the paper we focus on Kepler splittings, we consider in detail the eccentricity factor in the Wisdom-Holman errors seen in Fig. 2.

The Kepler problem is special in its degeneracies; it conserves...
all orbital elements, except for the mean longitude $\lambda$, which advances linearly. This implies that when alternating Kepler steps with perturbation steps, only perturbation components that change the particles’ semi-major axes (which determine the mean motions $d\lambda/dt$) do not commute.

For example, for a perturbation that only changed the eccentricity vector’s magnitude and direction, it would not matter whether the Kepler or perturbation step acted first, since the Kepler step leaves those quantities unchanged. The ordering only matters whether the Kepler or perturbation step acted first, since the Kepler step advances linearly. This implies that when alternating Kepler steps with perturbation steps, only perturbation components that change the semi-major axes (if the perturbation acts first, the mean motion will change, and when the Kepler step operates, the particle will end up at a different mean longitude than if the Kepler step had acted first with the original value of the mean motion.

We can then see that for any purely radial force like the oblateness perturbation, the change in orbital energy (and thus semi-major axis) $F \cdot v$ vanishes for circular motion. If we slowly increase the initial eccentricity, the perturbation will add and extract major axis (and thus semi-major axis) energy as each body moves toward and away from pericentre (every time unit) and apocentre (every half-time unit). We note that the flat errors for $e < 10^{-5}$ in the bottom plot of Fig. 2 are an artefact of our numerical setup.

The above example highlights the power of splitting methods’ geometrical properties. Because all the error terms in the modified differential equations (e.g., Eq. 7) consist of nested commutators of $L_d$ and $L_B$, any symmetries (or near-symmetries) are inherited in the integrator’s error properties to all orders. This means any higher-order splitting scheme that tries to correct for higher and higher order terms in the BCH expansion (e.g., Wisdom et al. 1996; Laskar & Robutel 2001) will always have such a geometrical suppression in its leading error term.

This is not typically emphasized in the literature on symplectic N-body integration, presumably because the Kepler and interaction steps do not have any such symmetries (there are always conjunctions between planets that change the semi-major axes at leading order). However, many astrophysically important effects are highly symmetric potentials, e.g. multipole gravitational potentials, radiation pressure, simple general relativistic corrections, etc. This fact, combined with the highly degenerate Kepler problem, often lead to strongly suppressed errors for the nearly circular or nearly coplanar orbits we often want to model.

In summary, splitting methods are often powerful not only for their long-term conservation of important quantities, but also because they can yield orders of magnitude higher accuracy for a fixed timestep. In the above case, a specialized second-order scheme split into separate Kepler and perturbation steps reduced the errors by a factor of $\epsilon_{1/2}^2 = 10^{-5}$ over short timescales compared to what one would obtain with a generic second-order Runge-Kutta scheme, even more over long timescales.

### 3.3 Dissipative Forces

While dissipative systems are qualitatively different from the conservative cases above, we now show that we can similarly understand their error behaviour using the above framework. As should be clear from the operator-centred development throughout the paper, we could take any differential equations, split them in any way we please, and if we could find time-evolution operators for those subproblems, then we could compose them and find their error behavior through the BCH formula. In particular, the error estimates

---

7 Actually it would not completely vanish in this problem due to the much weaker perturbations from the additional planet that would at these points become the dominant effect.

8 The physical argument is correct for all geometric eccentricities (i.e., the value measured by looking at the shape of the physical orbit in space, which includes the effects of the perturbation) but the typical subtlety arises when dealing with osculating eccentricities (i.e. the transformation from positions and velocities to unperturbed 2-body Kepler orbits, which ignores the effects of the perturbation) smaller than the size of the perturbation $e$. Because for convenience we initialize our orbits using osculating elements, decreasing the initial osculating eccentricities below $\epsilon_{1/2}$ does not make them any more circular in geometric space, because the oblateness perturbations kick them away from Keplerian motion at order $\epsilon_{1/2}$. So in our integrations, the geometric eccentricity that goes into the argument above reaches a flat floor at $\sim \epsilon_{1/2}^2$, with a corresponding error $\epsilon_{1/2}^2 h/T^2 \sim \epsilon_{1/2}^2 h/T^2$, which matches the flat regime on the left of the bottom panel of Fig. 2.
we developed in Sec. 2.4 and applied in Sec. 3.1 come from the modified differential equations, without reference to conservative or dissipative forces.

On the other hand, one important distinction for symplectic systems is that the fact that the composition of two Hamiltonian operators must be Hamiltonian implies that all the BCH error terms are also Hamiltonian (Sec. 2.2). By contrast, when the perturbation is dissipative, the presence of $L_\mu$ operators in the error terms of the modified differential equations (e.g., Eq. 7) shows that there is additional damping (or injection of energy) introduced by the splitting scheme. This generically leads to secular drifts, as we show in the following example.

Consider a single planet in orbit around its primary with an initial orbital eccentricity of 0.1 and an orbital period of 1 year, subject to a simple damping force directed opposite to the planet’s velocity vector $\mathbf{v}$,

$$\mathbf{F} = -m \frac{\mathbf{v}}{2r_\mu},$$

where $m$ is the planetary mass. This parametrized prescription orbit-averages to yield inward migration with the semi-major axis decaying exponentially on an e-folding timescale $\tau_e$ (Papaloizou & Larwood 2000). We set $\tau_e = 1000$ years, and integrate for three damping timescales $\tau_e$. In this time, the planet’s semi-major axis moves inward by a factor of $\exp(3) \approx 20$, and the orbital period decreases by a factor of $\exp(9/2) \approx 90$.

We integrate the system using a second-order Kepler splitting $K(h/2) \circ D(h) \circ K(h/2)$, where $K$ evolves the planet on a Kepler orbit, and $D$ damps the motion according to Eq. 14. We adopt a timestep of $10^{-3}$ times the planet’s initial orbital period, and choose to approximate $D(h)$ by integrating across the timestep using a fourth-order Runge-Kutta scheme (see Sec. 5).

With dissipation, we no longer have a conserved quantity to track. Instead, as a proxy for the exact propagator $P$, we also integrate the system with IAS15 (Rein & Spiegel 2015), a high-order adaptive-timestep method whose accuracy reaches machine precision. We plot in orange in Fig. 3 the relative error between the energy calculated in our Kepler splitting integration and the one with IAS15.

One can see a clear secular drift. While the errors in the symplectic evolution in Fig. 2 oscillate and average out to yield flat time evolution, dissipative errors systematically overdamp or undersmooth.

This scaling can be understood in the same way we analyzed the conservative case. The second-order splitting scheme error estimate is (Sec. 2.4) $\varepsilon_2(h/T)^2$, with $T$ the planet’s orbital period, and $\varepsilon_2 = T/\tau_e$. However, like in the case of $J_2$ perturbations discussed above, there is additionally a geometric suppression of the error by a factor of the eccentricity

$$\frac{\varepsilon_2(h/T)^2}{e}.$$  

To see this we have to go one step further than the arguments in Sec. 3.2. There we argued that only the components of the perturbation that change the particles’ semi-major axes don’t commute with the Kepler step. Here the force is always pointed exactly opposite the particle’s velocity, so the two steps definitely do not commute. But now we additionally have to ask whether the order of operations matters for the quantity whose error we are measuring, i.e., the energy. For a circular orbit, it would not. Whether or not the Kepler step moves the planet along the circle before or after the step does not affect the energy loss, since the problem is azimuthally symmetric. So again the energy errors are suppressed by a factor of $e$, though in this case there would be no suppression of the phase errors. We consider phase errors more carefully in Rein et al. (2019a).
4 HIGHER ACCURACY SPLITTING METHODS AT FIXED COMPUTATIONAL COST

4.1 Symplectic Correctors

Ignoring any improvements resulting from any symmetries in the problem (Sec. 3.2), the splitting errors for a second-order Wisdom-Holman scheme should oscillate on orbital timescales with an amplitude of $\sim \epsilon (h/T)^{10}$. However, these high frequency error oscillations should be unimportant to the long-term evolution through the averaging principle (Wisdom & Holman 1991).

Wisdom et al. (1996) go further and show that these high-frequency oscillations of $O(\epsilon)$ can be efficiently removed through a near-identity canonical transformation, which they term symplectic correctors, from the real action-angle variables to mapping variables. In this picture, the short term oscillations intuitively arise from a mismatch in initial conditions. When going from the real system to the modified mapping, one has to correct the initial conditions to modified 'mapping coordinates'.

An integration is then of the form

$$C^{-1}(h) \circ S_{ABA}(h) \circ C(h), \quad (16)$$

where each $S_{ABA}(h)$ represents a timestep of the second-order Wisdom-Holman map (Eq. 6), and $C(h)$ and $C^{-1}(h)$ are the symplectic corrector transformations to and from mapping variables at the beginning and end of the integration. Wisdom et al. (1996) furthermore prove that this canonical transformation removes all the error terms in the modified differential equations\footnote{We note that ordering matters once we specialize to systems where one operator is dominant. Wisdom et al. (1996) (with corrections in Wisdom 2006) give correctors for the $\mathcal{A}(h/2) \circ \mathcal{B}(h) \circ \mathcal{A}(h/2)$ second-order scheme. The correctors for the alternative second-order scheme $\epsilon \mathcal{B}(h/2) \circ \mathcal{A}(h) \circ \epsilon \mathcal{B}(h/2)$ would have different coefficients.} for the Wisdom-Holman map (Eq. 7) of $O(\epsilon h^2)$. In particular, we take the same problem of a single planet migrating inward from Sec. 3.3, except we shorten the semi-major axis damping timescale $\tau_d$ to 100 (initial) orbital periods, and integrate for one damping timescale.

We briefly note that these correctors for the second-order Wisdom-Holman map $\mathcal{S}_{ABA}$ (Eq. 6) can also be used for the first-order scheme $\mathcal{SAB}$ (Eq. 3) with a simple modification. As far as the first-order scheme is concerned, the second-order scheme is a reasonable approximation to the exact propagator $\mathcal{P}$, so

$$\mathcal{P}(h) \approx \mathcal{A}(h/2) \circ \mathcal{B}(h) \circ \mathcal{A}(h/2) \circ \mathcal{A}(-h/2) \circ \mathcal{B}(h) \circ \mathcal{A}(h/2),$$

and thus $C_1(h) = \mathcal{A}(h/2)$ acts as a first-order corrector for the first-order splitting scheme to yield $\mathcal{S}_{ABA}$. Then, if desired, one could use a higher order correctors $C_n$ in Eq. 16. This trick of combining half steps of the second-order scheme in the middle of the integration is used widely (in REBOUND this is referred to as synchronization), but rarely described as a corrector that can be extended to higher order\footnote{Jack Wisdom realized this long before we did (personal communication).}.

Finally, the fact that correctors compensate for the leading errors provides a straightforward way to interpret what the scheme is getting wrong. For example, inverting Eq. 17 yields

$$\mathcal{S}_{ABA}(h) \approx \mathcal{A}(h/2) \circ \mathcal{P}(h) \circ \mathcal{A}(-h/2), \quad (18)$$

i.e., to leading order, the first-order scheme $\mathcal{A}(h) \circ \mathcal{B}(h)$ is equivalent to performing an exact step along the true solution $\mathcal{P}(h)$, except with the mistake of taking additional forward and backward Kepler half-steps before and after the fact.

4.2 Symplectic Correctors Are Splitting Correctors

While Wisdom et al. (1996) derived symplectic correctors through physically motivated canonical perturbation theory specific to Hamiltonian systems, they point out that the formulas could also have been derived solely from the Lie algebra (commutators) of the two operators using Lie series and the BCH formula. But as mentioned above, the BCH formula is a statement about the composition of non-commutative operators and has nothing to do with symplecticity.

‘Symplectic’ correctors therefore should correct any second-order splitting scheme involving a dominant operator and a perturbation. We now demonstrate explicitly that this is true even for dissipative perturbations. ‘Symplectic’ correctors are therefore more widely applicable than their name implies. Specifically, they represent ‘weak splitting correctors.’ In fact, like Wisdom et al. (1996), we rediscovered that similar ideas were applied as far back as Butcher (1969) to general Runge-Kutta methods with no restriction to Hamiltonian systems.

In particular, we take the same problem of a single planet migrating inward from Sec. 3.3, except we shorten the semi-major axis damping timescale $\tau_d$ to 100 (initial) orbital periods, and integrate for one damping timescale.

We consider three integration schemes. First we apply the first order scheme $\mathcal{K}(h) \circ \mathcal{B}(h)$. From Secs. 2.4 and 3.3, this will yield errors $\sim \epsilon_0 (h/T) \epsilon$. We then run a separate integration using first-order correctors (Eq. 17), whose error should scale as $\epsilon_0 (h/T)^2 \epsilon$, and an integration with third-order correctors $C_3$ (Eq. 16), which should scale as $\epsilon_0 (h/T)^3 \epsilon$ for small $h$. The results are shown in Fig. 4, with all curves following the above estimates (black dashed lines).

Several features are apparent from Fig. 4. First, when thought of as a ‘symplectic’ corrector, it might seem surprising that symplectic correctors applied at the beginning and end of the integration (when the energy has dissipated by a factor of 3 in between) would give a more accurate result. But as argued above, in our case there is nothing specifically symplectic about them, they are simply weak splitting correctors. Second, while all the timesteps considered are small compared to the initial orbital period of 1 year, it is the final orbital period (vertical dashed line) that matters. We see that for timesteps $\gtrsim 10\%$ of the final orbital period, the error rises dramatically, a consequence of higher order terms in the BCH series becoming important.

It might seem counterintuitive that symplectic correctors can improve integrations involving dissipation. In particular, one might
There are ways to address this using time regularization. As discussed in the text, ‘symplectic correctors’ fix the leading scheme errors to provide the expected scaling. As discussed above, this is not an issue for position-dependent forces, since they can be trivially incorporated into the interaction step, Eq. 10. In that case Eq. 11 remains the exact analytical solution, using the accumulated accelerations from all the position-dependent effects.

However, many important astrophysical effects like the migration forces considered in the previous section, post-Newtonian corrections, tides, etc., are velocity-dependent. Previous authors have proposed incorporating such velocity-dependent perturbations in either the Kepler step (Malhotra 1994), in the interaction step (Cordeiro et al. 1996), or as a separate step (Touma & Wisdom 1994). The methods achieve comparable accuracy (Cordeiro et al. 1996), so we focus on the methods of Touma & Wisdom (1994) and Cordeiro et al. (1996), which fit directly into the framework discussed above.

In this case, the velocity-dependent accelerations vary across the perturbation step as the velocities change, so Eq. 11 is no longer exact; it is now merely an Euler step approximation, accurate only to first order in $h$. In this section we consider the errors that such an Euler step introduces, and compare it to higher-order numerical approximations to the evolution across the perturbation step in this general case where the propagator cannot be found analytically.

### 5.1 Hamiltonian Velocity-Dependent Forces

We first consider Hamiltonian velocity-dependent effects, which we will find are qualitatively different from ones involving dissipation. We note that while we will find such perturbations cause severe numerical problems for these previously proposed methods, those authors were interested in dissipative perturbations, which we consider in Sec. 3.3.

A good test case is the velocity-dependent first-order post-Newtonian correction for general relativity (Anderson et al. 1975), where, given the dominant central mass in planetary applications, we ignore second-order corrections of order the planet-star mass ratio. The equations of motion can be derived from a Hamiltonian whose conservation we can use to track the numerical accuracy.

In all integrations we adopt the first-order splitting

$$\mathcal{K}(h) \circ \mathcal{G}(h),$$

where the corresponding differential equations are

$$\mathcal{K}z : v_i = v_i,$$

$$\mathcal{G}z : v_i = a_{Kep}^i(r),$$

$$\mathcal{G}rz : v_i = 0,$$

$$v_i = a_{GR}^i(r, v).$$

with the $a_{GR}^i$ given in Eq. B5.

We note that it is possible to find a splitting for this post-Newtonian Hamiltonian for which the evolution under each operator can be solved analytically (Saha & Tremaine 1994). We nevertheless choose to use the above splitting both to explore the effects of imperfect approximations across the perturbation step, and with a view toward making the REBOUNDx library a general-purpose tool for integration. Such Hamiltonian velocity-dependent perturbations change the relationship between the particles’ physical velocities.

---

12 There are ways to address this using time regularization.
and their momenta through Hamilton’s equations, so this relationship varies depending on what forces are added to an integration. To minimize the necessary logic and possible pitfalls, we choose \textsc{reboundx} to always use the splitting \textit{20}, which can be applied to both conservative and dissipative velocity-dependent forces. This is the same setup as in both \cite{ToumaWisdom94} and \cite{Cordeiro96a}; we now compare their choice of integrating across the GR step using a first-order Euler approximation (Eq. 11) to using higher-order methods.

We first note that in this case of Hamiltonian velocity-dependent forces, \textcite{Mikkola98} proposes using a low-order symplectic scheme like the implicit midpoint method to salvage the symplecticity of the scheme. However, even a symplectic scheme will not be symplectic when applied the non-canonical position and momentum variables of Eq. 20, like all the above authors consider\textsuperscript{13}. Therefore, one might not actually expect any advantage from using a symplectic scheme across the perturbation step.

We test this empirically on the K2-137 system (\cite{Smith2017}), for which post-Newtonian effects are important. It consists of a 0.89 Earth radii ($R_\oplus$) planet on a 4.3 hour ($\omega_1 \approx 0.01$ AU) orbit around a 0.46 solar-mass primary. We assign the planet the mass of the Earth. While in reality tides should have circularized the orbit, we inflate the initial orbital eccentricity to 0.01, in order to see the apsidal precession induced by the relativistic effects. The post-Newtonian perturbation strength is $\epsilon_{\text{GR}} = 3GM/m^2 \approx 3 \times 10^{-6}$ (e.g., \cite{Nobili86}).

We compare integrating across the GR perturbation step with the second-order, symplectic implicit midpoint method advocated by \cite{Mikkola98} to three non-symplectic schemes: first-order Euler (\cite{ToumaWisdom94, Cordeiro96a}), second-order Ralston’s Runge Kutta (RK2), and fourth-order Runge Kutta (RK4). In all cases the timestep is $\approx 8.1\%$ of the innermost planet’s orbital period. This implies a relative error of $\sim \epsilon_{\text{GR}}(h/T)e \sim 2 \times 10^{-9}$ (black arrow).

In the RK2, RK4 and implicit midpoint integrations in Fig. 5, the truncation errors remain much smaller than the splitting scheme errors, and the simulations are visually indistinguishable, showing flat, oscillatory errors. As argued above, applying a symplectic algorithm like the implicit midpoint method to non-canonical variables does not yield a symplectic scheme, so none of these integrations are exactly symplectic. Nevertheless, by ensuring that the truncation errors across the perturbation step are much smaller than the splitting scheme errors, we only see the oscillatory symplectic behavior.

Of course, the most salient feature in Fig. 5 is that the error in the Euler method of \cite{ToumaWisdom94} and \cite{Cordeiro96a} (blue) error grows secularly, and in fact the planet is sparsely ejected after approximately 5.8 billion orbits, or in under 3 million years. This behaviour too can be analytically estimated.\textsuperscript{13}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Integration of the ultrashort period planet K2-137b under relativistic effects with an inflated orbital eccentricity of 0.01. Naive incorporation of the post-Newtonian accelerations into the Wisdom-Holman scheme (Euler method, blue) leads to a secular error growth, and the innermost planet escapes the system within 6 billion orbits or less than 3 million years. Integrating across the perturbation step using higher order methods, whether they are symplectic (implicit midpoint) or not (RK2, RK4), all give oscillatory errors that remain constant over the length of the integration.}
\end{figure}

5.2 Truncation Errors

For general perturbations that can not be integrated analytically, the local truncation error across the perturbation timestep must be incorporated into the error budget.

The local truncation error $\Delta z_{\text{pert}}^{(n)}$ across a single timestep using a one-step, order-$n$ method comes from the first neglected term at order $n + 1$ in a Taylor series approximation to the solution $z = f(t)$ (we focus on a single scalar coordinate $z$ for simplicity),

$$\Delta z_{\text{pert}}^{(n)} = \frac{f^{n+1}(\xi)}{(n+1)!} h^{n+1},$$

where $f^{n+1}$ is the $n + 1$th derivative of the true solution across the perturbation step, and $\xi$ is an unspecified time in the range $[t, t + h]$. If we assume a solution $f(t) = \exp(2\pi it/T_{\text{pert}})$, with the characteristic perturbation timescale $T_{\text{pert}}$ a factor of $\epsilon$ longer than the orbital timescale $T$, we have

$$\Delta z_{\text{pert}}^{(n)} = \frac{(2\pi)^{n+1} \epsilon^{n+1}}{(n+1)!} \left( \frac{h}{T} \right)^{n+1}.$$

This is the error incurred every timestep. We can obtain the worst-case relative global truncation error after an integration time $t$ by assuming the local errors add coherently, and multiplying Eq. 22 by the number of steps $t/h$,

$$\frac{\Delta z_{\text{pert}}^{(n)}}{z} \sim \frac{(2\pi)^{n+1} \epsilon^{n+1}}{(n+1)!} \left( \frac{h}{T} \right)^{n+1} \left( \frac{t}{T} \right).$$

We overplot this estimate for the Euler method of \cite{ToumaWisdom94} and \cite{Cordeiro96a} ($n = 1$) in Fig. 5 and see it gives a good match to the numerical behaviour. The truncation errors for the second-order RK2 method would be lower by a factor of $\epsilon_{\text{GR}}(h/T) \sim 10^{-7}$ and thus never become visible.

We might, however, expect different behavior under weak dissipation. To reach Eq. 23 we assumed the worst-case scenario that the one-step errors added coherently. At a more detailed level, one must consider how such errors are propagated by the dynamical flow itself. Consider a set of trajectories in phase space within a differential error volume around the true trajectory. By the divergence theorem, the fractional growth of this volume element across

\textsuperscript{13} As a simple analogy, consider a canonical (and thus symplectic) transformation that rotates 2D Cartesian coordinates. If we instead consider a transformation that first converts to non-canonical polar coordinates, naively applies the same symplectic rotation matrix to the polar coordinates, and then converts back to Cartesian coordinates, the sequence of transformations will not be canonical.
a timestep is given by the product of $h$ and the divergence of the vector field of trajectories $V \cdot \nabla$. In a Hamiltonian system like above, this divergence vanishes through Hamilton’s equations (this is Liouville’s theorem), so, at best, the dynamics are neutral for the accumulation of errors\textsuperscript{14}. By contrast, under weak dissipation, the divergence of the flow is $O(\epsilon)$ and negative, so errors contract by $O(\epsilon h)$ each timestep as the dynamics brings nearby trajectories together. One might therefore expect that integrating across the perturbation step with even a first-order Euler method would be good enough, given that one adds one-step errors $\sim O(\epsilon^2 h^2)$ (Eq. 22) slower than they are damped by the dynamics. We test this empirically in the following section.

5.3 Dissipation

Taking a somewhat realistic damping problem, we initialize two Jupiter-mass planets in a 3:2 mean-motion resonance around a Sun-like star, with the innermost planet initially at 0.90 AU using celmech\textsuperscript{15}. We then apply an exponential eccentricity damping force prescription (Papaloizou & Larwood 2000) to both planets with an $e$-folding decay timescale of $\tau_e \approx 1000$ initial inner-planet orbits using the modify_orbits\_forces implementation in REBOUNDx. The outer planet is then additionally acted on by an inward migration force (Eq. 14) with a timescale 100 times longer $\tau_m \approx 0.1$ Myr. The eccentricities grow as the system evolves deeper into resonance until the migration forces pushing the system into resonance are balanced against the dissipative divergence of the orbits caused by the eccentricity damping (e.g., Batygin & Morbidelli 2012; Lithwick & Wu 2012; Goldreich & Schlichting 2014), reaching equilibrium eccentricities of 0.022 and 0.025 for the inner and outer planet, respectively.

In Fig. 6 we then plot the energy errors during the subsequent integration, where the eccentricities remain in a quasi-steady state as the two plants migrate inward together, on a timescale approximately two times longer than the $\tau_e$ estimate above (two equal mass planets have to be moved). We use the second-order scheme\textsuperscript{16}

$$\mathcal{K}(h/2) \circ D(h/2) \circ J(h) \circ D(h/2) \circ \mathcal{K}(h/2),$$

and integrate across each damping step using the four methods of Sec. 5.1. We integrate for $4 \times 10^7$ initial orbits, which translates to approximately 4000 eccentricity-damping timescales, or 2 semimajor-axis damping timescales. We adopt a small fixed timestep of $10^{-4}$ of the innermost planet’s initial orbital period to accommodate the shrinking orbital periods, which decrease by a factor of $\approx 25$ over the integration. As above, we compare the energies to the energies in a ‘perfect’ integration with IAS15.

\textsuperscript{14} In reality, an initial parcel of trajectories within some error volume would get sheared out at fixed volume
\textsuperscript{15} celmech is a publicly available code for semi-analytic manipulations of the classical disturbing function in celestial mechanics. Among other things, it can perform transformations between orbital elements and resonant variables: https://github.com/shadden/celmech
\textsuperscript{16} We provide a simple argument why one can always use such a time-symmetric splitting with multiple operators to create a second-order scheme. The timestep coefficients for each operator must sum to unity in order to match the modified differential equations to the true ones at zeroth order, and by writing an explicitly time-symmetric scheme, we ensure that the additional terms in the modified differential equations cannot depend on the sign of $h$ (e.g., Eq. 7). The scheme is therefore second-order by construction.

In this case the scheme errors are dominated by the interplanetary, resonant interactions. As in Sec. 3.3, we see in Fig. 6 that the energy errors grow due to the shrinking orbital periods. Like in Fig. 5, the RK2, RK4 and implicit midpoint integrations are visually indistinguishable. By contrast, while the integrations using the first-order Euler method across the damping step have higher errors, it is not a runaway effect as in the conservative case in Fig. 5. As discussed in Sec. 5.2, this is because of the dissipative dynamics continually damping any accumulated energy errors.

In summary, the crudest, most computationally efficient Euler step across a velocity-dependent force step should give reliable results for dissipative forces, as advocated by Touma & Wisdom (1994) and Correia et al. (1996). However, such a scheme would yield qualitatively wrong results in long-term integrations with conservative velocity-dependent forces (Sec. 5.1). This can straightforwardly be solved by integrating across the perturbation step with a higher method; for high-accuracy explicit splitting schemes that go beyond this to correct even such higher-order integrations across the perturbation step, see Blanes et al. (2013).

6 REBOUNDx

In the previous sections we covered several technical aspects of how additional forces interact with splitting integration schemes like WHFast. However, many readers are likely interested in more practical questions of how to quickly get something working. We therefore implement the integration schemes and ideas discussed above in a new library, REBOUNDx, which can be found at https://github.com/dtamayo/reboundx.

REBOUNDx is library that provides tools and routines to accurately and self-consistently incorporate many different astrophysical effects to N-body simulations. It seamlessly interfaces with the REBOUND N-body integration package (Rein & Liu 2012). Like REBOUND, REBOUNDx is written in C, and provides a PYTHON wrapper to interface with other libraries. Also, like REBOUND, the REBOUNDx source code is machine independent. We implement a binary format to save REBOUNDx configurations that interfaces with...
the SimulationArchive class in REBOUND, making it possible to share and reproduce results bit by bit.

We have already implemented several common forces in REBOUNDx, some of which we used as examples in the discussion above. We hope to increase the number of supported astrophysical effects over time. REBOUNDx is an open-source project and we welcome the community to contribute new effects. We provide many tutorials and examples in the form of Jupyter notebooks that illustrate the library’s usage and how one can add new effects with minimal effort. We also share the scripts used to generate all the plots in this paper in a separate repository at https://github.com/dtamayo/reboundxpaper.

We now give an overview of the main concepts, structures and logic in the REBOUNDx package.

6.1 Which Integrator To Use?

In a scenario where all the forces (including interplanetary forces) are always small compared to the gravitational forces from the central body, we recommended using Wisdom-Holman Kepler splitting schemes. The implementation of this method in REBOUND is WHFast (Rein & Tamayo 2015a). This splitting will typically be fastest at reasonable levels of accuracy, and it provides strong conservation properties in long-term integrations.

Higher levels of accuracy are supported in cases where the forces can all be derived from position-dependent potentials by setting sim.riskfast.corrector = N, where N is the order of the corrector (Sec. 4.1). In velocity-dependent cases, correctors can be implemented manually (the notebook fig4.ipynb in this paper’s repository provides an example). For even higher levels of accuracy (without significant computational cost), one can use the WHCKL method (Rein et al. 2019b), which is an implementation of the kernel method of Wisdom et al. (1996) (see also Wisdom 2018) that also removes errors at O(ε²).

In cases with close encounters between planets, the motion will no longer be nearly Keplerian, and Kepler splittings will yield large errors. In cases where such close encounters are rare, hybrid integrators are powerful tools (Duncan et al. 1998; Chambers 1999; Rein et al. 2019c). In REBOUND, the MERCURIUS integrator retains efficiency by using a Kepler splitting when possible, and switches to an adaptive, high-accuracy integration with IAS15 for particles undergoing close encounters (Rein et al. 2019c). MERCURIUS supports additional forces, but only switches to IAS15 during close gravitational encounters. In the current implementation, any additional forces must remain small throughout the integration when MERCURIUS is used.

When the above conditions are not met, the safest option (and the default) in REBOUND is to use the IAS15 integrator (Rein & Spiegel 2015). Because it is an adaptive scheme, IAS15 can handle arbitrary forces and still yield solutions that are accurate close to the machine precision. IAS15 will typically be slower in cases where WHFast and MERCURIUS can be used. However, IAS15 will always be more accurate and is applicable to a wider variety of problems17.

6.2 REBOUNDx Structures

The two principal ways to incorporate additional effects into an N-body simulation are through forces and operators.

6.2.1 Forces

One way to implement an additional effect is to write it explicitly as a force. This force then contributes to the accelerations of all particles, in addition to the standard Newtonian gravity.

When the IAS15 integrator is used, the integrator automatically adapts the timestep such that integrations will be accurate close to machine-precision. If one of the various splitting integrators available within REBOUND (for example WH, MERCURIUS, and LEAPFROG) is used, then the additional forces are accumulated and applied in the integrator’s integration (or kick) step using Eq. 11. As discussed above, this is appropriate for position-dependent or dissipative forces (Secs. 3.1 and 5.3). For conservative, velocity-dependent forces, one should add a separate operator (Sec. 6.2.2).

For the hybrid symplectic MERCURIUS integrator (Rein et al. 2019c), only forces derivable from position-dependent potentials are supported.

6.2.2 Operators

Rather than calculating accelerations which are then integrated in the interaction or kick step of the integrator, operator steps yield solutions. Operators update the positions and velocities (or even masses or other particle parameters) to the value they should have at the end of a specified timestep.

Operators (e.g., \(D\)) together with a timestep (e.g., \(h/4\)) form a step (e.g., \(D(h/4)\)). The same operator can thus be used in multiple steps with different timesteps to construct higher-order splitting schemes (e.g., Eq. 24).

For cases where the solution for the positions and velocities at the end of the step is not known, or difficult to calculate analytically, the REBOUNDx object integrate_force operator can take a force, and integrate it across the timestep using any of the four integrators discussed in Sec. 5. The default is RK4. This is the recommended method for conservative, velocity-dependent forces (Sec. 5.1).

6.3 Ordering

All forces and operator steps are stored in linked lists such that they are executed in the reverse order they were added. Thus, to get a scheme consisting of the steps \(K(h) \circ D(h)\), the operator \(K\) should be added before \(D\), i.e. left to the right, since the last step acts first.

6.4 Timestep logic

REBOUNDx provides a simple API for adding effects into N-body integrations from the list of implemented options, demonstrated in the numerous examples in the documentation. However, it also provides lower-level functionality for customized tasks. The three main function pointers in a REBOUNDSimulation object that REBOUNDx sets and calls behind the scenes every timestep are

- **additional_forces**: This iterates through the linked list of attached forces to update particle accelerations.

---

17 By using adaptive timesteps, IAS15 can even be faster in cases, e.g., very eccentric orbits, where one has to choose a small timestep with WHFast or MERCURIUS in order to resolve a short timescale in the problem (e.g., pericenter passage).
• **pre timestep modifications.** This iterates through the linked list of attached operator steps to execute before the main REBOUND step.
• **post timestep modifications.** This iterates through the linked list of operator steps to execute after the main REBOUND step.

The user populates these lists by either adding forces and operators already implemented in REBOUNDx, or through user-defined functions in either PYTHON or C. By default, adding an operator to an integration adds half steps before and after the main REBOUND step to make a second-order scheme.

In Fig. 7, we schematically outline a full REBOUND timestep using the WHFast integrator. Different integrators switch out the reb_integrator_part1 and reb_integrator_part2 steps, but in all cases they rely on the gravitational and additional forces calculated in between them.

### 6.5 Coordinate systems

In contrast to many other N-body packages, REBOUND allows the user to set up the simulation in an arbitrary inertial frame. The coordinates are returned to the user in the same inertial frame.

The various splitting integrators in REBOUND use a number of different coordinates internally (e.g. Jacobi coordinates). In order to minimize logic and pitfalls, all REBOUNDx functions are always called after converting the particles’ positions and velocities back to Cartesian coordinates in the inertial frame. Force and operator implementations thus do not need to worry about indirect or non-inertial forces. This also implies that it is typically desirable to ensure that the net applied forces vanish. This is discussed in detail in Appendix C.

### 6.6 Parameters

Additional effects in an N-body simulation typically come with several associated parameters (e.g., a migration timescale for a damping force). REBOUNDx provides an interface which allows a user to attach an arbitrary numbers of parameters to forces, operators, and particles. The implementation is such that parameters remain tied to their parent structures even as objects get removed or reshuffled in the simulation.

### 6.7 Implemented effects

A current list of implemented effects can be found in the documentation. At the time of this writing, the following effects are available:

• Migration forces (Papaloizou & Larwood 2000)
• Eccentricity and inclination damping forces (Papaloizou & Larwood 2000)
• General relativistic 1 PN corrections (Newhall et al. 1983)
• General relativistic 1 PN corrections for a dominant central mass (Anderson et al. 1975)
• General relativistic 1 PN corrections with a simplified potential (Nobili & Roxburgh 1986)
• Radiation Forces (Burns et al. 1979)
• Mass loss
• Precession from conservative equilibrium tides (Hut 1981)
• General central forces
• Higher order gravitational moments, J2 and J4

![Figure 7. Schematic outline of a full REBOUND step when using the WHFast integrator. Functions which are part of REBOUNDx and called via function pointers are shaded.](image-url)
tive systems and can thus be more generally thought of as ‘weak splitting correctors’ (Sec. 4.2).

We also explored the general case where individual operator steps can’t be solved analytically, and one is forced to integrate across the perturbation step. We showed that the previously advocated approaches of incorporating additional forces into the Wisdom-Holman map (Touma & Wisdom 1994; Malhotra 1994; Cordeiro et al. 1996) work well for dissipative effects (Sec. 5.3), but give qualitatively wrong answers for conservative, velocity-dependent forces like post-Newtonian corrections of general relativity (Sec. 5.1).

Finally, we described REBOUNDx, an open-source C library for incorporating additional effects into REBOUND N-body integrations, together with a convenient PYTHON wrapper. Users can either choose to add an already implemented astrophysical force, or easily implement new effects. We hope that this modular, open-source framework for robustly incorporating new effects into a variety of N-body integration schemes will help the community to run more accurate and realistic simulations. We encourage others to contribute to this library of astrophysical effects.

ACKNOWLEDGMENTS

We would like to thank John Chambers for an insightful and constructive review that greatly improved this manuscript. We would also like to thank Jack Wisdom and Scott Tremaine for helpful discussions. Support for this work was provided by NASA through the NASA Hubble Fellowship 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. This research has been supported by the NSERC Discovery Grant RGPIN-2014-04553 and the Centre for Planetary Sciences at the University of Toronto. Support for this work was provided by NASA through contract NAS5-26555. This research was made possible by the open-source projects Jupyter (Kluyver et al. 2016), IPython (Pérez & Granger 2007), and matplotlib (Hunter 2007; Droettboom et al. 2016).

REFERENCES

Anderson J. D., Esposito P. B., Martin W., Thornton C. L., Muhleman D. O., 1975, ApJ, 200, 221
Batygin K., Morbidelli A., 2012, The Astronomical Journal, 145, 1
Benítez F., Gallardo T., 2008, Celestial Mechanics and Dynamical Astronomy, 101, 289
Blanes S., Casas F., Farrés A., Laskar J., Makazaga J., Murua A., 2013, Applied Numerical Mathematics, 68, 58
Brown E. W., 1931, Monthly Notices of the Royal Astronomical Society, 92, 104
Burns J. A., Lamy P. L., Soter S., 1979, Icarus, 40, 1
Butcher J., 1969, in Conference on the numerical solution of differential equations. pp 133–139
Chambers J. E., 1999, Monthly Notices of the Royal Astronomical Society, 304, 793
Cordeiro R., Gomes R., Martins R. V., 1996, Celestial Mechanics and Dynamical Astronomy, 65, 407
Dunlap J., 1992, Richmond: Willman-Bell,— c1992, 2nd ed.
Droettboom M., et al., 2016, matplotlib, matplotlib v1.5.1, doi:10.5281/zenodo.44579, http://dx.doi.org/10.5281/zenodo.44579
Duncan M. J., Levison H. F., Lee M. H., 1998, AJ, 116, 2067
Forest E., Ruth R. D., 1990, Physica D: Nonlinear Phenomena, 43, 105
Fujii M., Iwasawa M., Funato Y., Makino J., 2007, Publ. Astron. Soc. Japan, 59, 1095
Galley C. R., 2013, Physical review letters, 110, 174301
Galley C. R., Tsang D., Stein L. C., 2014, arXiv preprint arXiv:1412.3082
Goldreich P., Schlichting H. E., 2014, AJ, 147, 32
Gröbner W., 1967, Die lie-reihen und ihre anwendungen. V ol. 3, Deutscher Verlag der Wissenschaften
Hairer E., Lubich C., Wanner G., 2006, Geometric numerical integration: structure-preserving algorithms for ordinary differential equations. Springer Science & Business Media
Hernandez D. M., 2019a, arXiv preprint arXiv:1904.03364
Hernandez D. M., 2019b, MNRAS, 486, 5231
Hernandez D. M., Bertschinger E., 2018, Monthly Notices of the Royal Astronomical Society, 475, 5570
Hernandez D. M., Dehnen W., 2017, Monthly Notices of the Royal Astronomical Society, 468, 2614
Hunter J. D., 2007, Computing In Science & Engineering, 9, 90
Hut P., 1981, Astronomy and Astrophysics, 99, 126
Kinoshita H., Yoshida H., Nakai H., 1990, Celestial Mechanics and Dynamical Astronomy, 50, 59
Kluyver T., et al., 2016, Positioning and Power in Academic Publishing: Players, Agents and Agendas. p. 87
Laskar J., 1989, Nature, 338, 237
Laskar J., Gastineau M., 2009, Nature, 459, 817
Laskar J., Robutel P., 2001, Celestial Mechanics and Dynamical Astronomy, 80, 39
Lithwick Y., Wu Y., 2012, The Astrophysical Journal Letters, 756, L11
Machin J., 1737, Philosophical Transactions of the Royal Society of London, 40, 205
Malhotra R., 1994, Celestial Mechanics and Dynamical Astronomy, 60, 373
Mikkola S., 1998, Celestial Mechanics and Dynamical Astronomy, 68, 249
Mikkola S., Innenk K., 1999, Celestial Mechanics and Dynamical Astronomy, 74, 59
Mikkola S., Tanikawa K., 1999, Celestial Mechanics and Dynamical Astronomy, 74, 287
Naoz S., Kocsis B., Loeb A., Yunes N., 2013, The Astrophysical Journal, 773, 187
Neri F., 1987, Dept. of Physics, University of Maryland
Newhall X. X., Standish E. M. J., Williams J. G., 1983, A&A, 125, 150
Newton I., 1687, Philosophiae Naturalis Principia Mathematica. Auctore Js. Newton. doi:10.3931/rara-440.
Nobili A. M., Roxburgh I. W., 1986, in Relativity in Celestial Mechanics and Astrometry. High Precision Dynamical Theories and Observational Verifications. p. 105
Papaloizou J., Larwood J., 2000, Monthly Notices of the Royal Astronomical Society, 315, 823
Pelleceppey F., Portegies Zwart S., 2012, Monthly Notices of the Royal Astronomical Society, 420, 1503
Pérez F., Granger B. E., 2007, Computing in Science and Engineering, 9, 21
Petit A. C., Laskar J., Bour G., Gastineau M., 2019, arXiv preprint arXiv:1905.13240
Plummer H., 1896, Monthly Notices of the Royal Astronomical Society, 56, 317
Portegies Zwart S., 2018, Science, 361, 979
Portegies Zwart S., McMillan S., 2018, Astrophysical Recipes; The art of AMUSE, by Portegies Zwart, Simon; McMillan, Steve. ISBN: 978-0-7503-1321-6. IOP ebooks. Bristol, UK: IOP Publishing, 2018
Preto M., Tremaine S., 1999, The Astronomical Journal, 118, 2532
Rambaut A. A., 1890, Monthly Notices of the Royal Astronomical Society, 50, 301
Rauch K. P., Holman M., 1999, The Astronomical Journal, 117, 1087
Rein H., Liu S.-F., 2012, Astronomy & Astrophysics, 537, A128
Rein H., Spiegel D. S., 2015, MNRAS, 446, 1424
Rein H., Tamayo D., 2015a, Monthly Notices of the Royal Astronomical Society, 452, 376
Rein H., Tamayo D., 2015b, MNRAS, 452, 376
Rein H., Tamayo D., 2019, Research Notes of the AAS, 3, 16

MNRS 000, 000–000 (0000)
APPENDIX A: OPERATOR THEORY

While not strictly necessary for the development in the main text, we review here some of the gaps glossed over in Sec. 2. Using the same notation, we seek a solution to the differential equations

\[ \dot{z} = \mathcal{L}_F z, \]  

(A1)

where here we define the Lie derivative more carefully. The Lie derivative with respect to a set of differential equations \( z = \dot{F}z \) acts on functions \( g(z) \) to return their total derivative. Since throughout the paper we treat only cases that are explicitly time-independent, we have

\[ \mathcal{L}_F g(z) = \sum_i \frac{\partial g}{\partial z_i} \dot{z}_i, \]  

(A2)

where the \( \dot{z}_i \) are given by \( \dot{F}z \).

With this definition, we can Taylor expand the solution around the current time,

\[ z(t + h) \approx z(t) + h \frac{\partial z}{\partial t} \bigg|_{t} + \frac{h^2}{2!} \frac{\partial^2 z}{\partial t^2} \bigg|_{t} + \ldots \equiv \left[ 1 + h \mathcal{L}_F + \frac{h^2}{2!} \mathcal{L}_F^2 + \ldots \right] z(t). \]  

(A3)

As a concrete example, consider the simple one-dimensional differential equation \( \dot{z} = \dot{z} \). We have \( \mathcal{L}_F z = \dot{z} \), \( \mathcal{L}_F^2 z = \mathcal{L}_F \dot{z} = 2 \dot{z} \), etc. Plugging into Eq. A3 yields the Taylor expansion of the exact solution \( \dot{z}(t) = (1/(1 - h^2)) \).

By analogy to the Taylor expansion for the exponential, this is often written as\(^{16}\)

\[ z(t + h) \approx \exp(h \mathcal{L}_F) z(t) \equiv \mathcal{T}(h) z(t). \]  

(A4)

In words, \( \mathcal{T}(h) \) is a time evolution operator, i.e., an integrator, that updates the phase space variables across a timestep. We then proceed to split the differential equations into two pieces \( \dot{z} = \mathcal{L}_F z = (\mathcal{L}_B + \mathcal{L}_A) z \) as in the main text.

One subtle notational issue that has led to some errors in the literature is that the BCH formula is typically written as

\[ e^{h \mathcal{L}_A} e^{h \mathcal{L}_B} = e^{h \mathcal{L}_B} \]  

(A5)

with

\[ \mathcal{L}_S = \mathcal{L}_B + \mathcal{L}_A + \frac{h^2}{12} [\mathcal{L}_A, [\mathcal{L}_B, \mathcal{L}_A]] + O(h^3). \]  

(A6)

However, \( e^{h \mathcal{L}_A} e^{h \mathcal{L}_B} \) is not equivalent to splitting the scheme one might naively implement on a computer \( e^{h \mathcal{L}_A} \circ e^{h \mathcal{L}_B} \).

In \( e^{h \mathcal{L}_A} e^{h \mathcal{L}_B} \) \( z \), \( e^{h \mathcal{L}_A} \) acts on a function \( g(z) = e^{h \mathcal{L}_A} z \), and therefore introduces partial derivatives of \( g \) through Eq. A2. By contrast, the integrator \( e^{h \mathcal{L}_A} \circ e^{h \mathcal{L}_B} \) first evaluates an intermediate \( \dot{z} = e^{h \mathcal{L}_B} z \), and these updated values are used to yield a final answer of \( e^{h \mathcal{L}_A} \dot{z} \).

The BCH formula applies to the former sense (one can straightforwardly expand the three exponentials in Eq A5 to obtain the first terms in Eq. A6), but this interpretation is not particularly useful for writing integrators. Under iterated mappings, the repeated chain rules would quickly become unwieldy. By contrast, \( e^{h \mathcal{L}_A} \circ e^{h \mathcal{L}_B} \) is straightforward to implement on a computer, and fortunately can be simply related to the scheme in the BCH formula to analyze its error properties. This is sometimes referred to as the Verstaunghessatz (Gröbner 1967) (see Lemma 5.1 in Chapter III of Hairer et al. (2006)); we provide a simple derivation and example here.

Looking back at Eq. A3, one could just as well have Taylor expanded any scalar function \( f(z) \), and concluded in Eq. A4 that the time evolution operator \( e^{h \mathcal{L}_A} \) also evolves \( f(z) \) from time \( t \) to time \( t + h \) (with \( f(z) = z \) in Eq. A3 as a special case). Thus, since Eq. A5 is of the form \( e^{h \mathcal{L}_B} f(z) \), with \( f(z) = e^{h \mathcal{L}_A} z \), we must have

\[ e^{h \mathcal{L}_A} f(z) = f \circ (e^{h \mathcal{L}_A} z), \]  

(A7)

i.e., evolving \( f(z) \) across the timestep (LHS) is equivalent to first evolving \( z \) and applying the function \( f \) to the updated variables (RHS). Therefore,

\[ e^{h \mathcal{L}_A} \circ e^{h \mathcal{L}_B} = e^{h \mathcal{L}_B} e^{h \mathcal{L}_A}. \]  

(A8)

This is the reason for the sign flips at all odd orders in Eq. 5.

As one might expect, Eq. A8 has led to some confusion in notation and sign errors in the literature (e.g., footnote 3 in Saha & Tremaine 1992 and Eq. 10 in Hernandez & Dehnen 2017). However, one reason why this is probably not typically noticed or pointed out is that most integrators used in the astrophysics literature are time reversible, and thus will not have any sign flips in their corresponding BCH formulas given that they are invariant to flipping the order of operations (e.g., Yoshida 1993 and the detailed analysis in Hernandez & Dehnen 2017).

As an explicit example of the above, consider a non-dimensionalized simple harmonic oscillator with position \( x \), momentum \( v \), and \( z = (x, v) \). The Hamiltonian is \( H = \frac{v^2}{2} + \frac{x^2}{2} \), and we use a kinetic-potential splitting \( H_T = \frac{v^2}{2} \) and \( H_V = \frac{x^2}{2} \). The corresponding differential equations are given by Hamilton’s equations,

\[ \mathcal{L}_T z : \dot{x}_T = \frac{\partial H_T}{\partial v} = v, \quad \dot{v}_T = -\frac{\partial H_T}{\partial x} = 0, \]  

(A9)

\[ \mathcal{L}_V z : \dot{x}_V = \frac{\partial H_V}{\partial v} = 0, \quad \dot{v}_V = -\frac{\partial H_V}{\partial x} = -x. \]  

(A10)

(11)

\(^{16}\) Throughout the paper we assume that the differential equations do not depend explicitly on time. If they did, an integral would be required in the exponential of Eq. A4.
The propagators can easily be found by integrating their respective differential equations, or by noting $L^r T(z) = L^r T \cdot z = 0$ for $n \geq 2$, so

$$e^{iL^r T} z = (1 + h L^r T) z = \left(\frac{x + hv}{v} \right) .$$

(A12)

$$e^{iL^r T} z = (1 + h L^r V) z = \left(\frac{x}{v - hx}\right) .$$

(A13)

We first evaluate the composition used in the \textit{S/AB} integrator (Eq. 3), $e^{iL^r T} o e^{iL^r V} z$, where $e^{iL^r T} z'$ is evaluated at $z' = e^{iL^r T} z$

$$(e^{iL^r T} o e^{iL^r V}) z = \left(\frac{x + h(v - hx)}{v - hx}\right) .$$

(A14)

By Eq. A8, this should be equivalent to

$$e^{iL^r T} e^{iL^r V} z = (1 + h L^r V) \left(\frac{x + hv}{v}\right) = h \left(\frac{\tilde{v}_x + hv}{\tilde{v}_y}\right) .$$

(A15)

which indeed matches Eq. A14.

**APPENDIX B: FIRST ORDER POST-NEWTONIAN CORRECTIONS**

General relativistic corrections to Newtonian dynamics are detailed throughout the literature, but since we use these effects as one of the main tests for the various schemes in this paper, we present the equations and details of our implementation for completeness.

The first order post-Newtonian corrections (1PN) are $O(v^2/c^2)$, where $v$ is the characteristic orbital velocity and $c$ is the speed of light. We have implemented various levels of approximation. In \textit{gr}\_full, we include the full (1 PN) equations of motion, which are implicit and computationally expensive, but would be appropriate for equal-mass bodies.

For planets around a single star, one can typically ignore corrections to the GR perturbation of order the planet-star mass ratio. This significantly simplifies the equations, but involves velocity-dependent accelerations. This approximation is implemented in the \textit{gr} effect in \textit{REBOUNDx}, and is the case discussed in Sec. 5.

As a final level of approximation, one is typically interested in the apsidal precession effect from general relativity, in which case it is possible to write down a potential only depending on the particles’ position, which, in an orbit-averaged sense gives the correct precession rate (Nobili & Roxburgh 1986). This is implemented in \textit{gr\_potential}, and has the advantage that it can be integrated analytically (Sec. 3.1), and is therefore computationally cheapest. However, this approximation introduces errors in the phases and instantaneous elements of $O(v^2/c^2)$. Typically, \textit{gr\_potential} is sufficient for planetary applications around a single star since it captures the correct secular behavior, but both \textit{gr\_potential} and \textit{gr} are only appropriate for single-star systems. For higher multiplicity systems, one must use \textit{gr\_full}.

Naoz et al. (2013) provide the full Hamiltonian for a triple system (see also Schäfer 1987). The equations of motion are given by Newhall et al. (1983) and Benitez & Gallardo (2008). These accelerations and Hamiltonian are implemented in \textit{gr\_full}.

In the \textit{gr} implementation, we assume a dominant central mass, and ignore additional corrections of order the planet-star mass ratio. We outline our implementation for this effect in more detail since it differs from previous works, and is the subject of Sec. 5. The full Hamiltonian can be written as a sum of the Newtonian and 1PN Hamiltonians, $H = H_0 + H_\text{1PN}$, where

$$H_\text{1PN} = \sum_i \left\{ \frac{p_i^2}{2m_i} - \frac{1}{2} \sum_{ij} \frac{G m_i m_j}{r_{ij}} \right\} .$$

(B2)

Having already ignored terms of $O(m_i/m_0)$, the difference between barycentric, Jacobi or democratic heliocentric coordinates (and Jacobi or physical masses) in $H_\text{1PN}$ is negligible. Interpreting all positions and momenta as Jacobi coordinates considerably simplifies the equations of motion, since the kinetic term in $H_i$ remains a diagonal sum of the momenta. Application of Hamilton’s equations yields (e.g., Saha & Tremaine 1994),

$$\mathbf{r}_0 = \nabla_{\mathbf{p}_0} H = \frac{p_0}{m_0} = \mathbf{v}_0,$nabla_{\mathbf{p}_i} H = \frac{p_i}{m_i} + \nabla_{\mathbf{p}_i} H_\text{1PN} = \mathbf{v}_i (1 + A_i) .$$

(B3)

where $A_i = -\frac{1}{c^2} \left(\frac{\tilde{v}_x}{2} + \frac{3\mu}{r_i} \right)$

(B4)

and the $\mathbf{v}_i \equiv \mathbf{p}_i/m_i$ are pseudo-velocities not equal to the physical velocities $\mathbf{f}_i$. We note that Eq. B3 would have additional terms in democratic heliocentric coordinates due to kinetic cross terms in $H_i$. We then have

$$\mathbf{r}_0 = \mathbf{\tilde{v}}_0 = 0,$nabla_{\mathbf{p}_i} H = \mathbf{a}_i + \mathbf{\tilde{v}}_i (1 + A_i) + A_i \mathbf{\tilde{v}}_i$$

(B5)

where

$$\mathbf{a}_i = -\frac{1}{m_i} \nabla_{\mathbf{p}_i} H = \mathbf{a}_i + \mathbf{\tilde{v}}_i \left(\frac{3\mu \mathbf{\tilde{v}}_i}{r_i^3} - \frac{\mu^2}{r_i^5} \right).$$

(B6)

$\mathbf{\tilde{v}}_i$ are the net Newtonian accelerations on particle $i$, and $A_i = -\frac{1}{c^2} \left(\mathbf{\tilde{v}}_i \cdot \mathbf{\tilde{v}}_i - \frac{3\mu}{r_i} (r_i \cdot \mathbf{f}_i) \right)$.

(B7)

We see that the accelerations (Eq. B5) are explicitly velocity-dependent, and must be treated with care in symplectic schemes (Sec. 5). As expected, $\mathbf{r}_0 = 0$ and the centre of mass moves at constant speed. Equation B3 needs to be solved implicitly for the pseudo-velocities $\mathbf{\tilde{v}}_i$, but this is easily accomplished iteratively, since the 1PN approximation is only appropriate when the perturbation is weak, and the physical and pseudo-velocities are approximately equal. Operationally, if like in \textit{REBOUND}, the integrator does not work in Jacobi coordinates, one simply calculates the Newtonian accelerations in inertial coordinates, transforms to Jacobi coordinates, calculates the Jacobi accelerations $\mathbf{\tilde{f}}_i$ and transforms back to inertial coordinates. Rein & Tamayo (2015a) provide unbiased transformation algorithms between inertial and Jacobi coordinates.
barycentre. Such a simulation’s accuracy will continually deteriorate as the system moves away from the origin and the precision on the inter-particle separations degrades due to subtracting ever larger nearly equal numbers. These errors may be acceptable depending on the application, but can slow down schemes like IAS15 (Rein & Spiegel 2015) that try to reach machine precision as they fail to converge with progressively smaller timesteps.

It is therefore numerically desirable to self-consistently model all back-reactions from additional effects so that the net force vanishes. While this is trivial for forces between pairs of particles in the simulation, one does not always wish to fully model all components.

Consider a net force $F_D$ from a protoplanetary disk of mass $M_D$ on a planet. In response, the disk should feel an equal and opposite net force, but there is no disk “particle” in the simulation on which to apply it. This uncompensated force will yield an acceleration on the COM. However, the disk is tightly coupled to the central star gravitationally, suggesting that this back-reaction should be effectively communicated to the central body. More quantitatively, if we take a characteristic orbital radius for the disk $r_D$, then we can define a characteristic timescale for the back-reaction $\tau_{BR} \sim (r_D/a_D)^{1/2}$, where $a_D$ is the net acceleration of the disk $F_D/M_D$. At the same time, the disk is gravitationally coupled to the primary on the Keplerian orbital period $\tau_K = 2\pi (G M_P)^{-1/2} / r_D^3$.

If $\tau_K \ll \tau_{BR}$, the central body will respond adiabatically to the accelerations of the disk, and the back-reaction could instead be added to the central star.

In other cases it might be better motivated to add the accelerations to the centre of mass of all or a subset of the particles, e.g., back-reactions onto a circumbinary disk. This corresponds to applying an acceleration $a = -F/M_{\text{tot}}$ to the appropriate set of particles, where $M_{\text{tot}}$ is their total mass.

In the case of forces that only depend on particle positions, the above choices are straightforwardly reflected in the structure of the Hamiltonians that govern the dynamics. Being able to precisely calculate these Hamiltonians provides a practical check on the numerical accuracy of the implementation for conservative systems. A Hamiltonian that depends on the radial positions $r_i$ and $r_j$ of two particles, $H(r_i - r_j)$ exerts equal and opposite forces on the two bodies. Similarly, one can show that for a Hamiltonian dependent on the distance from the system’s barycentre $H(r_i - \mathbf{r}_{\text{COM}})$, where $\mathbf{r}_{\text{COM}} = M_{\text{tot}}^{-1} \sum_j m_j \mathbf{r}_j$, Hamilton’s equations dictate that if particle $i$ feels force $F_i = -\nabla_{r_i} H(r_i - \mathbf{r}_{\text{COM}})$, all particles (including $i$) should feel an additional acceleration $-\mathbf{a}_{\text{COM}} = -F_i/M_{\text{tot}}$. Finally, for a Hamiltonian with positions referenced to the centre of mass of a subset of particles, Hamilton’s equations require that only those particles feel the reverse acceleration $-F_i/M_{\text{sub}}$, where $M_{\text{sub}}$ is the mass of the subset of particles. In the case where each particle’s position is referenced to the centre of mass of all interior bodies, this corresponds to the familiar Jacobi coordinates.