An Introduction to Gravity in the Solar System

Marc A. Murison
U.S. Naval Observatory, Washington, DC
murison@usno.navy.mil

25 October, 2005

Abstract

We discuss the role of gravity in the formation of the Solar System and in the motions of the planets and other bodies.

Subject headings: celestial mechanics—gravitation—history and philosophy of astronomy—ISM: globules—methods: analytical—methods: numerical—minor planets, asteroids—planetary systems: formation—planetary systems: protoplanetary disks—planets and satellites: formation—planets and satellites: general—solar system: formation—stars: formation

The XML version of this document is available on the web at
http://www.alpheratz.net/murison/papers/SolarSystemGravity/SolarSystemGravity.xml

The PDF version of this document is available on the web at
http://www.alpheratz.net/murison/papers/SolarSystemGravity/SolarSystemGravity.pdf

Figure 1: The Orion Complex Molecular Cloud (*Photo: Robert Gendler*). Birthplace of stars and solar systems.

1 Introduction

"*The most beautiful thing we can experience is the mysterious. It is the source of all true art and science.*" —Albert Einstein

Gravity plays a crucial role over the entire history of our Solar System, from its very beginings as a condensation inside a Giant Molecular Cloud (GMC) (see Figures 1 and 2) right up to the present day. Unlike the other known forces in Nature, gravity is very weak but acts over very large length scales – indeed, even across the universe. The large scale of interaction between bodies is the key to a profound influence of gravity over all things large in the universe. Our purpose here is to briefly survey the role of gravity in the Solar System: from formation to how the planets, asteroids, comets, and natural satellites move today. Along the way we emphasize the cultural view of science and rational thought and how that is crucial to understanding our place in the universe.

In the next section, we quickly outline the formation of the Solar System, starting with a GMC. Then we sketch the highlights of how we currently think the planets and other bodies formed from this tremendous accumulation of gas and dust. In Section 3 we list important milestones of human thought regarding the world and the universe, as well as the gradual realization of the order of things and, at last, of how planetary motion behaves. We realize that observation must always take precedence over ideology and preconceived notions if the goal is to learn about the world. In Section 4 we review the basic classical notions of how gravity works – that is, the nuts and bolts one needs to describe and predict planetary motions. In Section 5 we make a brief attempt to say what gravity *is*, as opposed to what it *does*. We conclude in Section 6 with a select smorgasbord of current research topics concerning gravity and its effects in the Solar System (all of which are fascinating and very exciting, of course!).

2 Formation of the Solar System

2.1 We Are Made of "Star Stuff"

The formation of a solar system begins with large condensations of gas and dust within a giant molecular cloud (Figs. 1 and 2). These condensations, over the course of millions of years, slowly collapse from their own self-gravity. These large condensations fragment and collapse into smaller regions of higher density, which individually collapse to form stars. Thus are star clusters born. The bright region in the center of Figure 1 shows the very brightest and most massive stars in just such a newborn cluster. The tremendous radiation pressure from these massive new stars blows away the intervening gas and dust from the cluster. Sometimes the resulting bubble in the GMC is close enough to the outside edge of the molecular cloud to blow a hole in its side, and once in a while we here on Earth are positioned so that we can see into the resulting cavity.

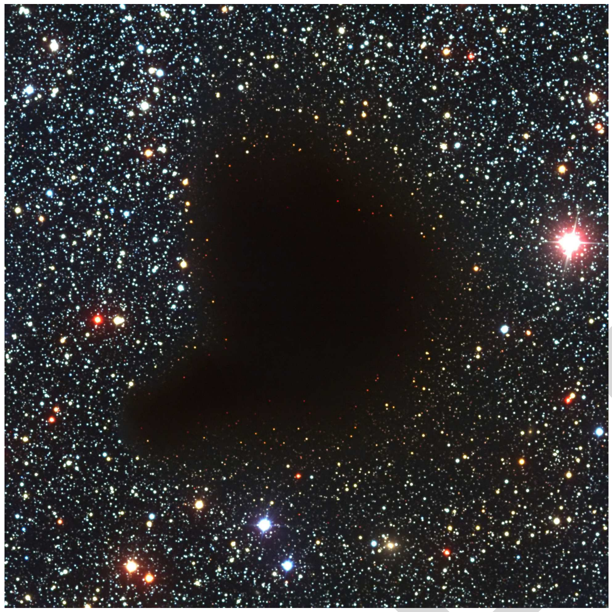
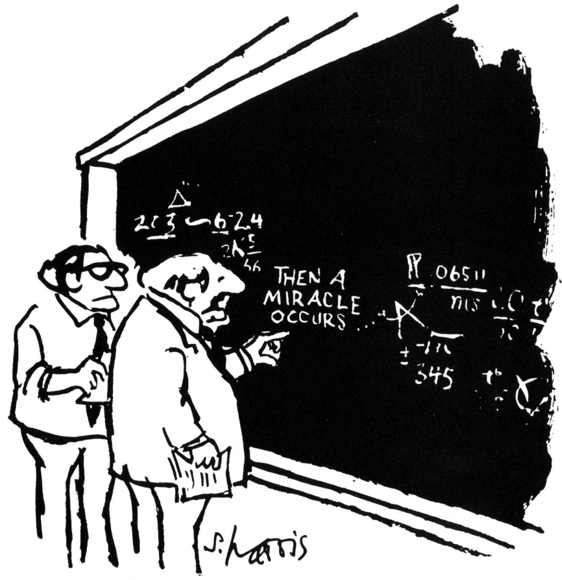


Figure 2: Barnard 68, an isolated dark globule (*Photo: European Southern Observatory*). Stars and their solar systems form when condensations inside globules of gas and dust such as this one collapse due to their own gravity.

2.2 A Miracle Occurs

The detailed physics of a collapsing interstellar cloud is not so difficult in the very begining stages of collapse (for a graduate student in physics or astronomy, anyway). That is, it is comparatively simple to show that the material of the cloud is subject to gravitational instabilities which will cause small density perturbations to grow into large, collapsing clumps within the GMC. Thus, starting with a giant molecular cloud, we think we understand the begining stages of gravitational collapse — the gas temperatures and pressures, the sizes of the nascent clumps, the motions of the gas, and how all these physical parameters interact. Gas and magnetic field pressures are negligable compared to the gravitational forces, since the dust and gas is both very massive and very cold. Call this stage A.

We also think we understand the later stages of collapse, wherein, at the centers of the smallest and densest of the condensations, temperatures and pressures reach the point at which nuclear fusion begins in what are now the cores of newborn stars — call this stage B. Our knowledge of subsequent stellar evolution is fairly well established for most types of stars, more or less.



"I think you should be more explicit here in step two."

However, it is not yet understood how the transition from stage A to stage B works in detail. We know it must happen, since we observe (via radio and infrared astronomy) giant molecular clouds containing interior clumps in the process of collapsing, and we observe newly-ignited young stars (called young stellar objects, also studied at radio and infrared wavelengths) embedded in more-evolved GMCs. In that in-between stage, gas pressure and magnetic field pressure — both of which act to counter the gravitational forces — are no longer ignorable and must be taken into account, yet the physical conditions inside these clouds is not well-enough known to allow us to do so.

Once we get past this stage of star formation and ignition, we then have the issue of forming planets out of the leftover material.

2.3 How do Planets Form?

We don't really know.

However, we do have some good ideas based on the

1. known physical processes,
2. educated guesses about the physical conditions that must have been present during the planet formation stage of the Solar System, and
3. sophisticated computer modeling

currently at our disposal. Planet formation is currently a very active area of research, and a more or less consistent picture is forming. The leftover material from formation of the central star is orbiting around the star, so it forms a flattened disk, called a protoplanetary disk (Figure 3). Starting with a disk of material circling a young central star, one can distinguish two, or perhaps three, regimes of subsequent planet formation: inner rocky planets, outer gas giants, and outer rocky/icy worlds.

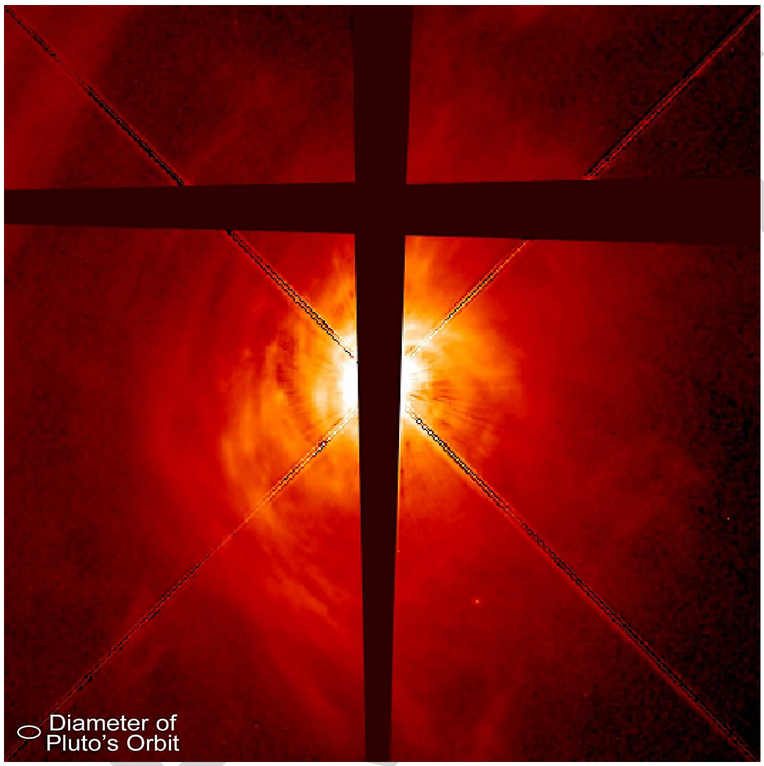


Figure 3: The inner part of the protoplanetary disk surrounding the star AB Aurigae. The dark cross is the shadow of an occulting bar that blocks most of the light from the central star. The diagonal lines are artifacts leftover from the removal of the diffraction spikes caused by the Hubble Space Telescope's secondary mirror supports. (*Photo: Space Telescope Science Institute*)

The inner solar system planets ("terrestrial planets") are thought to have formed by solid body accretion of kilometer-sized objects, which in turn were formed via collisional growth starting with micron-sized dust grains from a late-stage protoplanetary disk. Dust grains are present throughout the solar nebula as it collapses from the GMC. After the disk of material orbiting around the new Sun is formed, the dust grains, being heavier than the gas particles, sink towards the disk midplane and become concentrated (and therefore much more likely to suffer collisions with other dust grains). Figure 4 shows images of dust rings around two newly-formed stars. The timescale for settling to the disk midplane and growing to cm size particles is roughly 10^4-10^5 years at 1 AU. (1 AU = 1 *Astronomical Unit*

$= 150 \times 10^6$ km = the mean distance of Earth from the Sun's center.)

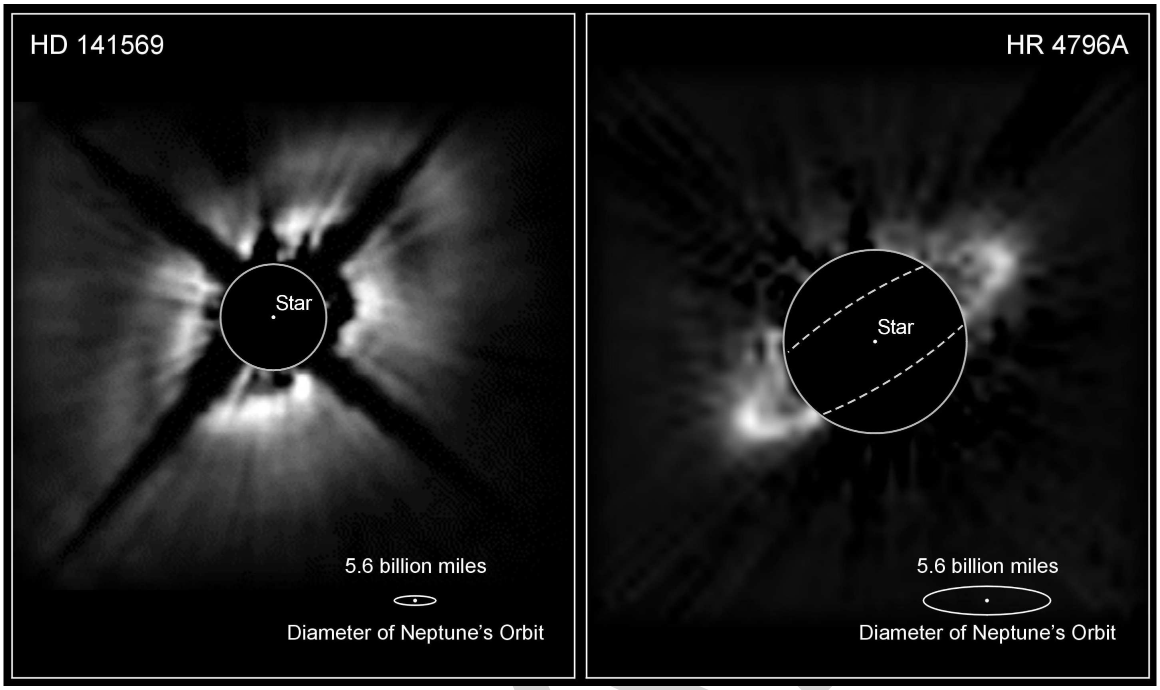


Figure 4: Rings of dust around the stars HD 141569 and HR 4796A. Most of the light from the central stars is blocked by an occulting bar. (*Photo: Space Telescope Science Institute*)

The processes by which dust grains in the protoplanetary disk collide and stick (or not) is not understood. One problem is that it is exceedingly difficult to perform laboratory experiments that might be able to reliably relate to the physical conditions of the early solar nebula and disk. We think we understand how the 0.1-1 micron size particles can collide and stick, eventually forming cm-sized pebbles. How to go from 1-10 cm up to 0.1-1 km size planetesimals is currently not known. Hence, the popular approach is to go at it from the other end: make educated guesses about the physical processes in computer models, let the models run, and see if reasonable-looking solar systems result. If not, then you've got the physics wrong; if so, then maybe you're on to something.

Once planetesimals of $\sim$10 km radius have formed, then mutual gravitational interactions, gas drag, disk-planet interactions, and accretion by collisions will soon lead to massive solid cores. The details of the interactions between the cores and the planetesimal swarm are very complicated and the subject of much activity. But we do think the cores somehow emerge from the process on stable orbits. How it all sorts out to get to that state is a bit fuzzy at present. In the inner solar system, these cores become the terrestrial planets, while in the outer Solar System they form the seeds for further growth into gas giants. There are still many hazy, or even completely unknown, parts to this story, but this is how a real solar system might get to the point of having a retinue of Earth-mass objects orbiting the nascent central star.

For the gas giants, there appear to be two possible formation scenarios. First, the planets could have formed via fragmentation of the protoplanetary disk and subsequent collapse of the fragments into planets. This cannot work for the inner solar system due to the higher temperatures and the higher radiation and particle wind from the Sun, preventing large blobs of material from collapsing before they're eaten away and dispersed. The disk fragmentation occurs because rotating disks of material are subject to various gravitational and fluid instabilities. The advantage of this model is that, once fragmentation begins, planets can form very quickly, which would be consistent with observations of extrasolar planetary systems. However, it is difficult to get a protoplanetary disk to fragment in such a way that planets could actually form. Condensation of fragments into planets requires axisymmetric disk fragmentation modes. Yet protoplanetary disks are relatively immune to axisymmetric instabilities and prefer asymmetric modes such as spiral waves, for which subsequent condensation is difficult.

The second outer planet formation scenario is accretion of gas and dust onto planetary cores. Again, the details are complicated and the subject of much research at present. One of the major problems is one of timing: at a certain point the cores become massive enough that runaway growth occurs. It is difficult to form planets quickly enough in this scenario, as the slow buildup before the runaway growth phase can last a very long time, posing a problem with modern observations. Thus, both modes of outer planet formation have their features and problems, and it is not clear which actually occurs or how in detail.

3 Of Human Flailings, or History, Shmistory!

Figure 5: *The School of Athens* by Raffaello Santi, 1509.

Deductive reasoning and the scientific method are relatively modern developments. Descartes (1596-1650) espoused, in the 1640s, what was then a new way of thinking (or, at least, it hadn't been summarized and clearly stated in print before then in the Western world): one should reject ideas based merely on *assumptions*, emotional beliefs, or *ideology*, and accept only those ideas which can be proved by or systematically deduced from *observation*. Deductive reasoning does make occasional appearances throughout recorded history, and it waxes and wanes on decadal time scales even in modern times. (Rather unfortunately, we appear currently to be at a low point in U.S. culture.)

This section, by way of short examples stated as theorems (for such they were at the time they were formulated), shows the thread of deductive (or not) reasoning through recorded Western history regarding the Earth, the Solar System, and the motions of the planets. These theorems are the fundamental precursors to the ability of the modern human race to understand gravity.

Theorem 1. *(Early Mesopotamia) The Earth is flat.*

Theorem 2. *(Ancient Greeks and others) Things go up. They come back down.*

Remark. These two theorems seem entirely self-evident. The first, of course, is quite wrong. It would not be until Isaac Newton before the mechanics of the second was understood, and that it, too, is only conditionally right. (Hint: rocket ship.)

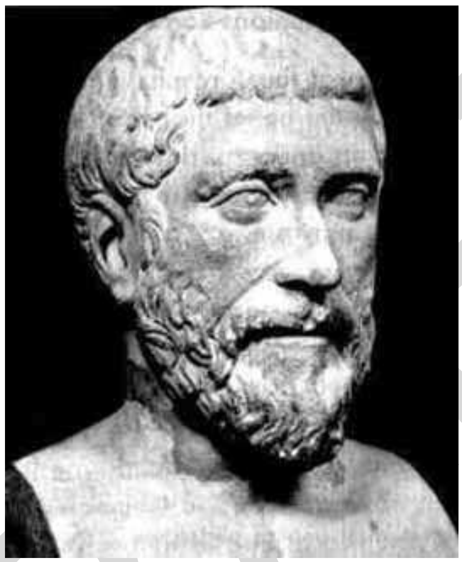


Figure 6: Pythagoras (ca. 569-500 BCE)

Theorem 3. *(Ancient Greeks around the time of Pythagoras, ca. 569-500 BCE) The Earth is round.*

Remark. Here we have early evidence of conclusions about the natural world based on deductive reasoning. Aristotle (384-322 BCE) argued that the Earth is a sphere because its shadow on the Moon during a lunar eclipse is round.

Theorem 4. *(Eratosthenes, 276-195 BCE) The Earth has radius roughly 4200 miles.*

Remark. Around 240 BCE, Eratosthenes *calculated* the radius of the Earth by *measuring* the lengths of shadows cast by a vertical stick at noon in two cities in Egypt (Syene and Alexandria) that differ in latitude. Eratosthenes was remarkably close: the Earth's radius is actually about 3963 miles. This is an early and classic example of what we now call the scientific method.

Remark. Eratosthenes, a Greek philosopher, was also head of the royal library at Alexandria, the greatest library of classical antiquity. Chalk one up for librarians.

Theorem 5. *(Aristotle, 384-322 BCE) The Earth is at the center of the Universe.*

Corollary 1. *The heavens, including the Sun and planets, revolve around the Earth.*

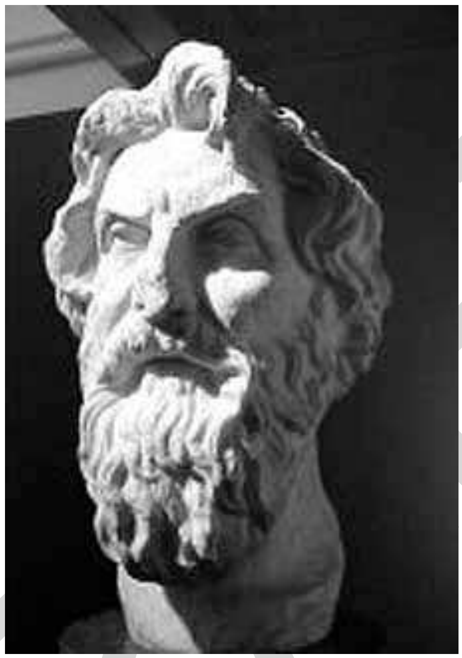


Figure 7: Aristarchus of Samos (310-230 BCE)

Theorem 6. *(Aristarchus, 310-230 BCE) The Sun is the center of the universe.*

Remark. Apparently, Aristarchus was the first to suggest that the Earth and planets revolve around the Sun, and that the Sun is the center of everything. Aristarchus is also famous for calculating the diameter of the Moon, based on his observations of the Earth's shadow during a lunar eclipse combined with the value of the diameter of the Earth from Eratosthenes.

Theorem 7. *(Hipparchus, ca. 190-120 BCE) The Earth's axis is not fixed in space.*

Remark. Hipparchus deduced from observations of star positions over time that the spin axis of the Earth changes direction. Later it was determined that the spin axis makes a circle around the ecliptic pole after about 25,770 years. This known as the precession of the equinoxes. Like all spinning tops being acted upon by outside forces, the Earth precesses. The forces causing the Earth to precess are the gravitational tugging of the Sun, Moon, and planets on the equatorial bulge of the Earth.

Theorem 8. *(Ptolemy, ca. 2nd century CE) The planets move on epicycles in their motions around the Earth.*

Remark. Here we have a classic example of muddled thinking. Belief had it that the Earth was the center of everything. Observations of the motions of the planets on the sky conflicted with this model. So Ptolemy tried to patch the model up by adding epicycles, rather than rethinking his assumptions and model. "Adding epicycles" is now a common metaphor for this kind of mistaken thinking. This is not to say that avoiding this pitfall is easy. Often we are faced with an inadequate model which mostly works but which we suspect may be fundamentally flawed, yet there is at the time no other recourse than to continue use of the flawed model while searching for a better one. However, using a flawed model with full knowledge that something is wrong is quite different from stubbornly sticking with mistaken assumptions in the face of contradictory observational evidence.

Theorem 9. *(Early Christian authors, ∼200-550 CE) The Earth is flat.*

Remark. Rational thought and observed fact can be lost, forgotten, or buried by ideology and ignorance.

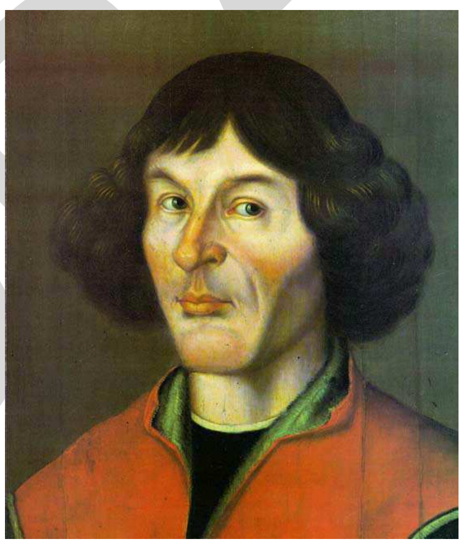


Figure 8: Nicolas Copernicus (1473-1543). Portrait from Toruń, beginning of the 16th century.

Theorem 10. *(Copernicus, 1473-1543) The Earth and the other planets move around the Sun.*

Remark. The scientific revolution of Western Europe begins.

Figure 9: Galileo Galilei, by Leoni (1624)

Theorem 11. *(Galileo, 1564–1642) Copernicus was right about the Solar System!*

Remark. Galileo observed that the Moon has mountains and craters, that Venus shows the full range of phases like the Moon, and that Jupiter has satellites. From his observations of the appearance and disappearance of the Jovian satellites, he concluded that these satellites are orbiting around Jupiter. This posed serious problems for the prevailing dogma that the Earth is the center of the universe. Galileo's observations of the phases of Venus proved that Venus orbits the Sun, as predicted by the heliocentric model of Copernicus.

Theorem 12. *(Galileo) Objects with different masses fall at the same velocity in a vacuum.*

Remark. Galileo insisted on performing rigorous experiments and looking for a mathematical description of the laws of nature. This approach set him apart from most "natural philosophers" that preceded him. It also put him in serious trouble with Inquisition of the Catholic Church.

"All he thinks about is that stupid ball."

4 What Does Gravity Do?

"*Nevertheless, it does move.*" —Galileo Galilei, 1633

By now we realize that this mysterious force we call gravity fundamentally affects the structure of and movements within the Solar System. With the advent of the Age of Enlightenment and the subsequent explosion of mathematics and curiosity about the natural world, we naturally turn to gravity and wonder how it works and what it is. The latter turns out to be an *exceedingly* deep question. If we can't quite discern its nature, then surely we can apply mathematics and observation in order to characterize what it does and how it makes moving objects behave. Employing mathematics and physical reasoning in combination, one can actually characterize the effects of this thing called gravity and even predict effects not yet observed. This new way of thinking reached its pinnacle with Isaac Newton and his laws of motion and gravity.

Newton's mathematical explorations and physical explanations are founded on the work of Kepler before him. (Indeed, it was Newton who famously said, in a letter to Robert Hooke in 1676, "If I have seen further it is by standing on the shoulders of Giants." This was probably meant as an insult, as Newton and Hooke disliked each other greatly, and Hooke was of very short stature. The origins of the quote are attributed to the young Roman poet Lucan (39-65 CE), in his epic poem *The Civil War*, "Pigmies placed on the shoulders of giants see more than the giants themselves." However, the ten surviving books of the (probably unfinished) *Civil War* do not contain the quote.) We leave aside for now the question of the *nature* of gravity and consider for a bit what it *does* – that is, how we can quantitatively characterise the basic *effects* of gravity.

Figure 10: Johannes Kepler (1571-1630)

4.1 Kepler the Mystic

Johannes Kepler (1571 - 1630), astrologer and Imperial Mathematician for Emperor Rudolph II in Prague, was a mystic and a mathematical prodigy with a keen interest in the beauty and harmony of the heavens. (Kepler also served as legal defense for his mother, who was arrested and tried as a witch. After fourteen months, during which torture was employed, she was finally set free on a legal technicality!) From observations of planetary positions obtained by his mentor, the meticulous and stubborn Danish astronomer Tycho Brahe (1546-1601), Kepler empirically deduced what became known as his three laws of planetary motion. He did not understand why the planets move the way they do, as the mathematics of the day was not advanced enough, but he nevertheless discovered from his tables of numbers certain mathematical relationships that planetary motions seemed to obey. He calculated the most exact astronomical tables known to that time. Indeed, the continued accuracy of the tables over several decades helped firmly establish the heliocentric theory of Copernicus (Theorem 10).

Theorem 13. *(Kepler's First Law) Planets orbit the sun in ellipses with the sun at one focus of the ellipse.*

Remark. The prevailing belief during Kepler's time was that the planets move on *circles* around the Sun. Kepler proved from the observational data that each planet actually moves on an *ellipse*, not a circle, with the Sun positioned at one focus of the ellipse, not the center. This was a triumph of observation and fact over belief and preconception. It was also an astounding feat of mathematical intuition on the part of Kepler.

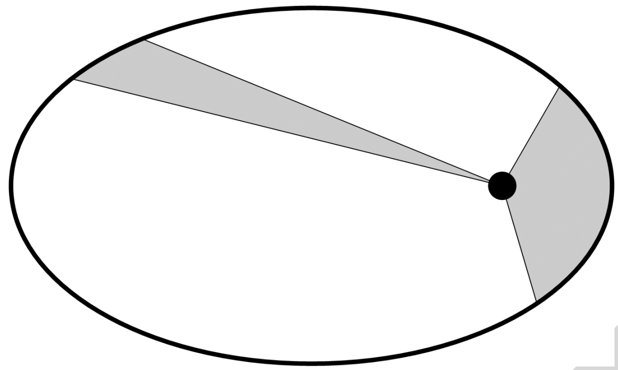


Figure 11: Kepler's Second Law: equal areas in equal time.

Theorem 14. *(Kepler's Second Law) Orbital motion sweeps out equal areas in equal times.*

Remark. This is a statement of what later became known as the conservation of angular momentum.

Theorem 15. *(Kepler's Third Law) The period P of an orbit squared is proportional to its mean distance a cubed.*

$$P^2 \propto a^3 \tag{1}$$

Remark. We now know that Kepler's Third Law is the result of the balance between the gravitational and centripital forces. The constant of proportionality involves the masses of the two bodies, M_1 and M_2, as well as Newton's constant of gravitation, G. Thus, the mathematical statement of Kepler's Third Law is more precisely written

$$P^2 = \frac{4\pi^2}{G(M_1 + M_2)} a^3 \tag{2}$$

This is easy to prove with circular orbits, but one needs calculus to show that eq. (2) is true for elliptical orbits as well, with a being the semimajor axis of the ellipse. In Kepler's day, the tools of classical mechanics and calculus had yet to be invented, so Kepler was only able to state the proportionality, eq. (1).

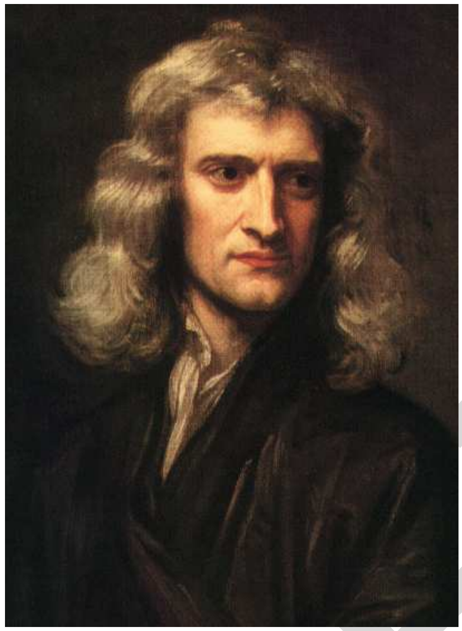


Figure 12: Isaac Newton, by Godfrey Kneller (1689)

4.2 Newton's Notions

Isaac Newton (1643-1727) invented differential calculus (as did Leibniz, independently) and developed the groundwork for all of classical mechanics. Kepler's three laws provided the basis for Newton's thinking on planetary motion and gravity. Newton's three laws of motion and his universal law of gravitation in turn provide the mathematical foundation for Kepler's three laws. It is difficult to overstate the importance of Newton's three laws of motion or of his development of what we have come to know as Newtonian gravity. They are the basis of all of classical mechanics.

4.2.1 The Three Laws of Motion

Theorem 16. *(Newton's First Law, the Law of Inertia) An object moves with constant velocity v unless acted upon by a net external force F.*

$$F = 0 \iff v = constant \tag{3}$$

Theorem 17. *(Newton's Second Law) A force F acting upon on object causes the object to accelerate. The acceleration a is proportional to the net applied force and inversely proportional to the object's mass m.*

$$F = m \cdot a \tag{4}$$

Remark. Since velocity is the change in distance with time, $v = \frac{dx}{dt}$, and since acceleration is the change in velocity with time, $a = \frac{dv}{dt}$, we have from eq. (4) the (vector) equations describing the motion of an object due to an applied force:

$$\begin{aligned} \frac{dv}{dt} &= \frac{1}{m} \cdot F(v, x, t) \\ \frac{dx}{dt} &= v \end{aligned} \tag{5}$$

The force may be a function of time, position, or velocity.

Remark. $F = ma$ is almost all of classical mechanics in a deceptively compact nutshell!

Theorem 18. *(Newton's Third Law) For every action there is an equal and opposite reaction.*

Remark. Suppose body A exerts a force on body B. We will denote this force by F_{AB}. Then

$$F_{AB} = -F_{BA} \tag{6}$$

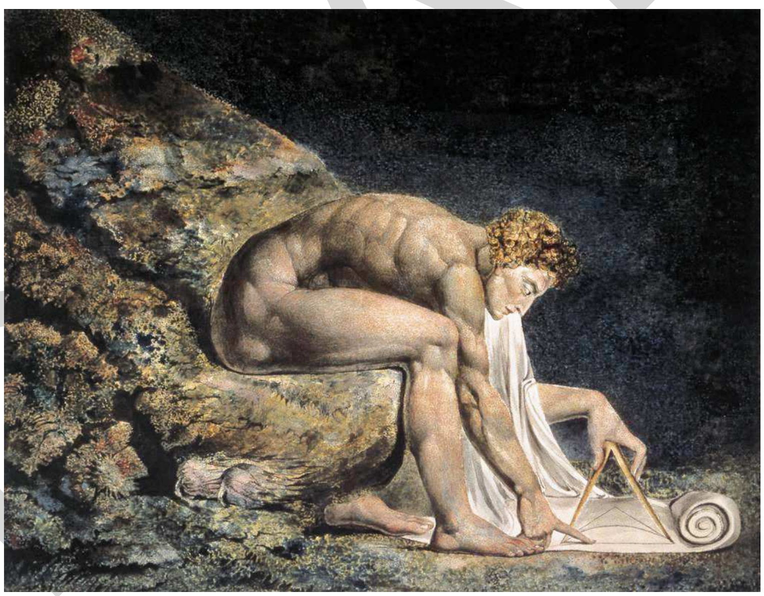


Figure 13: Isaac Newton, by William Blake (1795). Blake sought to exemplify the deeper significance of his philosophical thought in the tension between the immediate realism of his image and fantastic symbolism. Newton, man naked and created out of chaos, appears to be breaking through the chaos. He is discovering the law that is inherent in his own physical nature. Man has tasted of the fruit of the tree of knowledge, and now his intellect reveals to his astonished gaze the abstract reality of creation.

4.2.2 The Universal Law of Gravity

"*Gravity: it's not just a good idea, it's the law.*" —bumper sticker popular with nerds

Theorem 19. *(Newtonian Gravitation) Two masses M_1 and M_2 attract each other with a force that is proportional to their masses and inversely proportional to the distance r between them squared:*

$$F = G\frac{M_1 M_2}{r^2} \tag{7}$$

Remark. This law is deceptively simple in appearance. Yet it adequately covers almost (but not quite!) all that we know and observe regarding the motions of bodies in the Solar System. One begins to appreciate the complexity hidden in eq. (7) by considering the equations of motion that result for a body being influenced by the Newtonian gravity of $N-1$ other perfectly spherical bodies (the so-called *N-body problem*):

$$\frac{d^2\vec{\boldsymbol{r}}}{dt^2} = -G\sum_{k=1}^{N-1}\frac{M_k}{\left|\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}_k\right|^3}\left(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}_k\right) \tag{8}$$

There is a similar equation for each of the other bodies. The presence of each body is felt by all of the others. Furthermore, this is a system of nonlinear equations. These properties render them extremely difficult to solve, even for simplified ideal situations. However, it is just these properties that give rise to the most interesting phenomena in dynamical systems, such as chaotic motion.

The two-body problem can be solved analytically, meaning we can write down the complete description of the motions of the bodies using simple (or, more accurately, closed-form) mathematical functions. For example, the solution for the distance between the bodies in the two-body problem is given by the equation for an ellipse:

$$r = \frac{a\left(1 - e^2\right)}{1 + e\cos\theta} \tag{9}$$

where e is the eccentricity and θ is the azimuthal angle. This is what Kepler discerned from his tables of position observations as the basic motion of a planet around the Sun. But add just one more body and the problem cannot be solved analytically, except for certain highly-contrived circumstances. This wasn't proved until the brilliant French mathematician and theoretical astronomer Henri Poincaré (1854-1912) came along and did it near the end of the 19^{th} century. Astronomers and mathematicians have spent over three centuries working out the motions of the Solar System bodies using Newtonian gravity, and we are still nowhere close to a complete description.

4.3 Circular Motion

[...]

Definition 1. *Centripetal acceleration is the acceleration towards the center of a circular path that keeps an object in constant circular motion. Its magnitude is*

$$a_{centripetal} = \frac{v_c^2}{r} \tag{10}$$

where v_c is the magnitude of the velocity along the circular path.

[...]

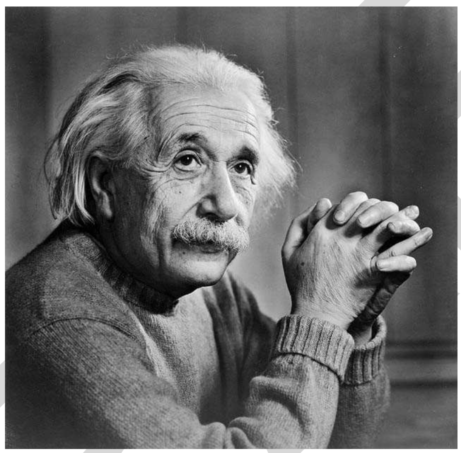


Figure 14: Albert Einstein (1879-1955) (*Photo: Yousef Karsh, 1948*)

5 What IS Gravity?

5.1 Einstein and Geometry

"*Reality is merely an illusion, albeit a very persistent one.*" —Albert Einstein

The concept of gravity is tenuous and ephemeral at best for most of us (including scientists!). After Newton, the concept of a force field developed. One can imagine a field of force emanating from an object – an electric field in the case of charged particles, or a gravitational field in case of massive bodies. There are just four known types of force in our universe (the strong and weak nuclear forces being the other two). The force field is rooted in the object as part of its very nature, and other objects interact with the field. This interaction produces accelerations, and we call the effect a force. If this sounds like a highly-theoretical construct, it is.

One of the serious problems with the concept of a force field in Newtonian gravity is that it requires action at a distance, a term meaning the instantaneous propagation of force. For example, the Newtonian equations of gravity make no allowance for the propagation of a disturbance (say, the movement of one planet affecting another some distance away) at a finite speed, which surely must be unphysical, since everything we see, touch, and measure involves finite speeds. (The speed of light is very large, but even it, too, is finite.) For example, the disturbances in an electromagnetic field propagate at the speed of light. We make practical use of them as radio waves. We are able in most cases to put up with this disturbing flaw in Newtonian gravity, since the propagation speed of gravitational disturbances, though finite, is still very large, so that in most applications it can be approximated as infinite. The resulting errors are extremely small.

However, towards the end of the 19^{th} century, technological improvements in measuring ability began to let us see inconsistencies between observation and theory. For example, the ellipse that Mercury moves along precesses (Figure 15). Now, the influences of the other planets will cause an orbit to precess, and we can calculate the magnitude of the effect. It turns out that there is an observed extra amount of precession that just could not be explained with existing theory. Newton was in trouble.

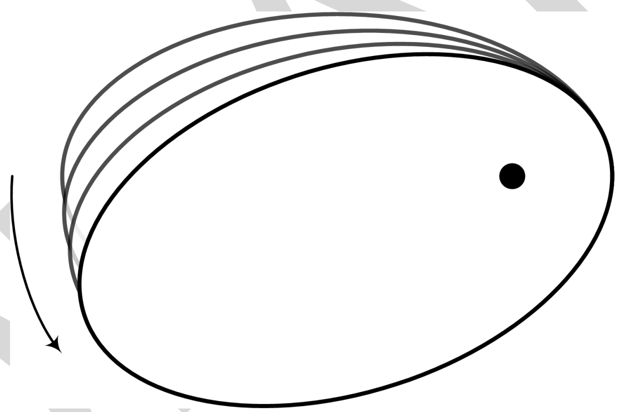


Figure 15: The slow precession of an elliptical orbit.

"*Matter and energy tell space (and space-time) how to curve. Space tells matter how to move.*" −John A. Wheeler, 1990

Albert Einstein (1879-1955) spent a decade developing his theory of General Relativity, finally publishing it in 1915. Einstein's great thought was to *geometrize* our concept of gravity: consider the warping of space and time caused by mass, and how that warp, or curvature, of space-time causes bodies to move.

[...]

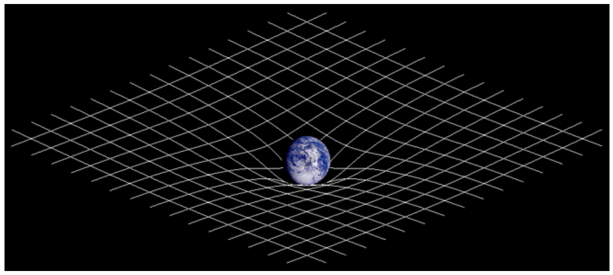


Figure 16: Illustrating the curvature of space-time via a 2-dimensional analogy. Mass alters the geometry of space-time, and the resulting curvature is gravity.

Einstein's field equations for general relativity and be written in tensor form as

$$\underbrace{G_{\mu\nu}}_{\text{Einstein tensor}} + \underbrace{\Lambda}_{\text{cosmological constant}} \cdot \underbrace{g_{\mu\nu}}_{\text{spacetime metric}} = \frac{8\pi G}{c^4} \cdot \underbrace{T_{\mu\nu}}_{\text{stress} - \text{energy tensor}} \tag{11}$$

where

$$G_{\mu\nu} = \underbrace{R_{\mu\nu}}_{\text{Ricci curvature}} - \frac{1}{2} g_{\mu\nu} \cdot \underbrace{R}_{\text{scalar curvature}} \tag{12}$$

is the Einstein tensor. (Think of a tensor, in this case, as a two-dimensional matrix.) The stress-energy tensor contains the information about a body's mass and energy, while the Einstein tensor contains information about the curvature of space and time. It is a deceptively simple equation, yet the concepts that it embodies, and the mathematics that it requires in order to solve even the most simple cases, are mind-boggling. Even so, it simplifies to Newtonian gravity for not-huge masses moving at slow speeds, such as are normally found in our Solar System.

[...]

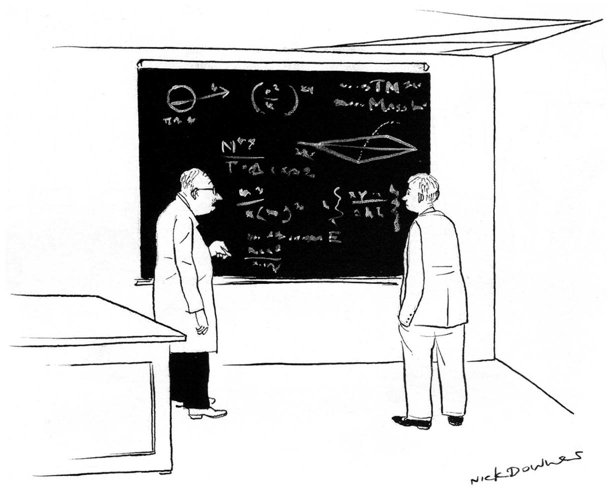
http://en.wikipedia.org/wiki/Ricci_tensor

http://en.wikipedia.org/wiki/Metric_tensor_%28general_relativity%29

5.2 Of Strings and Other Wiggly Things

http://en.wikipedia.org/wiki/Quantum_gravity

http://en.wikipedia.org/wiki/String_theory



"In layman's terms? I'm afraid I don't know any layman's terms."

6 Of Tintinnabulation and Children Swinging, or the Music of 1-Spheres and Resonances

"*What a tale of terror, now, their turbulency tells!*" —Edgar Allan Poe, 1849

In this section, we introduce some modern work being done in Solar System dynamics. The concept of resonance is a recurring and underlying theme, so first we will discuss what we mean by that in the context of Solar System motions. Then we consider the simplest type of resonance, called a mean-motion resonance, between two objects as they orbit about the Sun. Secular and spin-orbit resonances are two further categories of resonance, the first between two planets again, and the second a special type involving the spin of an object interacting with its own orbit. We discuss those next. Finally, we consider the next hierarchy of mean-motion resonances: resonances *inside* resonances. This is a sure-fire way of creating chaotic motions among the bodies involved.

[Okay, so what is a 1-sphere? It's mathspeak for circle (more properly a closed curve that can be smoothly deformed into a circle), which is a 1-dimensional sphere. A 0-sphere is a point, a 2-sphere is what we usually think of as a "sphere", a 3-sphere is a sphere in four-dimensional space, and so on.]

6.1 What is a Resonance?

Simply put, a resonance between any two coupled oscillatory phenomena occurs when the *periods* of oscillation are an integer ratio. (It is equivalent to require that their *frequencies* of oscillation are an integer ratio.) The basic idea of a resonance is familiar to anyone who has pushed a child on a swing. The swing, if pushed and then left alone, has a certain period of oscillation, or back-and-forth motion. Let's call this period P. If you give the swing a small push at intervals equal or close to P, then you will be pumping energy into the swing's motion. It is quite surprising how the accumulation of very small pushes will soon result in a very large amplitude of swinging. This phenomenon is resonance, and it is the same concept no matter how complex the physics and mathematics.

Another example is that of a singer breaking a crystal wine glass. (Yes, it can be done.) The wine glass has a natural frequency of oscillation, which we call the resonant frequency. If the singer can sing a note at that resonant frequency, the sound waves transfer small amounts of energy to the motion of the glass vibrations. In the same way you can get a child going very high on a swing with a succession of small pushes, the wine glass vibrations become very large after a succession of small pushes from the sound waves from the singer's voice. After a short time, the glass cannot handle the stresses of vibration and it shatters.

6.2 Mean-Motion Resonances and the Asteroids

In planetary motion, a mean-motion resonance is when the orbital periods of two orbits are a ratio of integers. For example, Pluto is in a 2:3 resonance with Neptune, completing two orbits of the Sun in the time it takes Neptune to orbit three times. Thus, at periodic intervals, Neptune and Pluto tug each other very slightly at the same locations in space. It turns out that resonances are of fundamental importance because of this periodic tugging. Resonances can cause either stability or instability of the associated dynamics, depending on the stability of the associated periodic orbit.

At a stable resonance, small perturbations will not change the dynamics. The Pluto-Neptune resonance is an example. The 2:3 resonance ensures that, whenever Pluto is closest to the Sun, Neptune is far away from Pluto, thus guaranteeing they will never have a close encounter. Another example is the case of the orbits of Jupiter's moons Ganymede, Europa, and Io: they are in a 1:2:4 orbital resonance. The critical angle involving their orbital longitudes, $\sigma = L_{\mathrm{Io}} - 3L_{\mathrm{Europa}} + 2L_{\mathrm{Ganymede}}$, librates around $\sigma = \pi$. (The critical angle will be explained below, in section 6.2.2.) Hence their mean motions are such that $n_{\mathrm{Io}} - 3n_{\mathrm{Europa}} + 2n_{\mathrm{Ganymede}} = 0$.

At an unstable resonance, small perturbations will grow, and the bodies will over time be driven away from their resonant configuration. This can lead to some pretty strange behavior. For example, certain resonances associated with the precession of Saturn's orbit cause asteroids located at certain points in the asteroid belt to be driven towards the inner solar system. This is where the supply of near-earth asteroids originates.

6.2.1 Mean-Motion Resonances in the Main Belt

The effects of mean-motion resonances with Jupiter can be seen in the distribution of the asteroids. [...]

- clearing and gathering due to resonances (asteroid belt)

- three-body problem and rotating reference frame
- some orbit resonance examples from the inner solar system
- resonance widths and chaos in the asteroid belt
- how to detect resonant motion
  - resonant angles
  - number theory and continued fractions

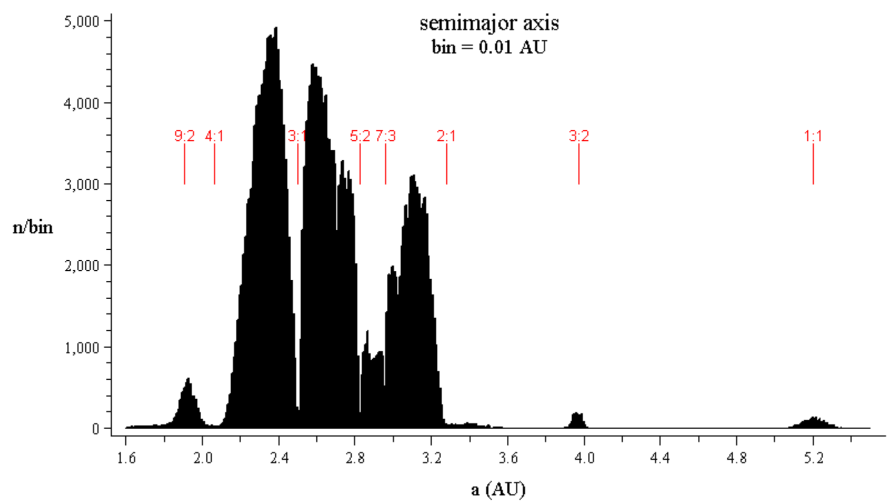


Figure 17: Histogram of semimajor axes of 298,000 asteroids. Major Jovian mean-motion resonance locations are marked in red. The Trojan asteroids are in 1:1 resonance near 5.2 AU, and the Hilda family of asteroids is stabilized by the 3:2 resonance near 3.9 AU.

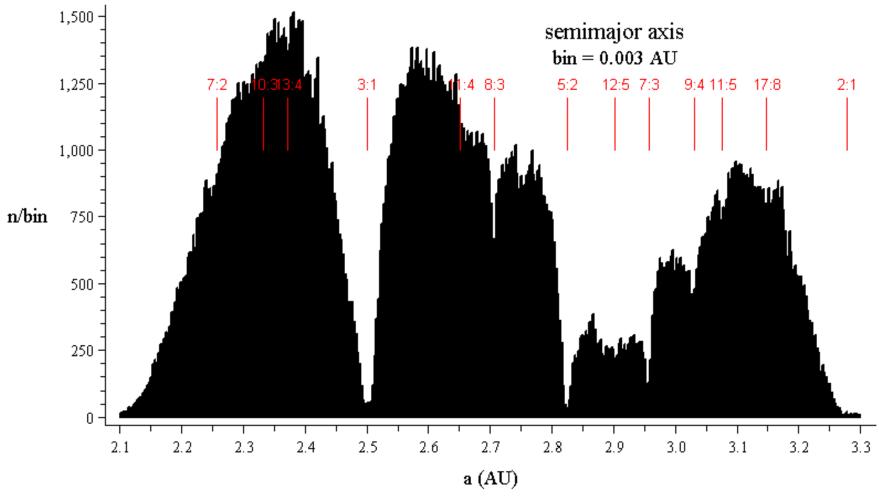


Figure 18: Semimajor axis histogram in the asteroid main belt. Several Jovian mean-motion resonance locations are marked in red. Clearly, the original distribution has been eaten away over the eons by the 2:1, 3:1, 5:2, 7:3, 9:4, 8:3, and even the 7:2, 11:5, and 17:8(!) resonances with Jupiter.

6.2.2 A Few Resonance Orbit Examples

In certain situations it is considerably simpler to study an orbit by switching to a rotating coordinate frame. Suppose we have the Sun, a planet, and an asteroid. We can make our coordinate system rotate with the planet so that the Sun is fixed at the origin and the planet is fixed at a spot on the X axis. This is the so-called rotating reference frame of the restricted three-body problem, and it is very handy.

Let us briefly mention a way to sort resonances according to their importance, or the strength of their effects onthe dynamics. A mean-motion resonance occurs when there is a $m : n$ commensurability between the mean motions of two orbiting bodies. We may write this as $(p + q) : p$, which for various reasons turns out to be a more useful form. The order of the resonance is then q, which is a measure of the importance, or perturbing influence, of the resonance. For example, we call a 3:2 resonance, where $q = 1$, is a *first-order* resonance. It is a stronger, more important resonance than, say, a third-order 7:4 resonance, where $q = 3$. Thus, the higher the order of a resonance, the weaker its effects.

Now back to the rotating frame. Figure 19 shows the orbit of Pluto in a frame that rotates with the planet Neptune. Here, the Sun is at the coordinate origin, and Neptune (not shown) is fixed at the crosshairs on the X axis at $x = 1$. The numerical integration spanned 100,000 years. Notice that in this frame Pluto executes a loopy path with two loops, one above and the other below the X axis. The double-loop pattern rocks back and forth over time, but the loops never extend any further than shown in the figure. This rocking to and fro is called *libration* (as opposed to *circulation*, where the pattern would make complete revolutions). Libration is a hallmark of resonance. If you can find in the motion of an orbit an integer combination of angles that librates instead of circulating, then you've found a resonance. This combination of angles is what we call a *critical angle*. In the case of $(p+q):p$ mean-motion resonances, we can define the critical angle σ as

$$\sigma = (p+q)\cdot L_{\text{ref}} - p \cdot L - q \cdot (\omega + \Omega) \tag{13}$$

where L_{ref} and L are the orbital longitudes (i.e., the azimuthal angles) of the reference body (Neptune in the case of Figure 19) and the body of interest (Pluto). The term involving the argument of pericenter and the longitude of the ascending node, $q \cdot (\omega + \Omega)$, is just a correction factor that makes things simpler in a mathematical sense. The important point to note is that different combinations of p and q can be applied to a numerical integration according eq. (13) and, if libration occurs, you've found a resonance.

Figure 20 shows two views of the critical angle for the case of the Pluto-Neptune resonance shown in Figure 19. Figures 21 and 22 show the case of a 3:5 resonance of the Trans-Neptunian Object 1994 JS with Neptune. This resonance is more complicated and, at second order, is weaker than the one Pluto is in, but as with Pluto we see that the resonance protects 1994 JS from encounters with Neptune. Our final example, shown in Figures 23 and 24, is a third-order, 4:7 resonance between the TNO 1997 CV_{29} and Neptune. It, too, is a stable resonance.

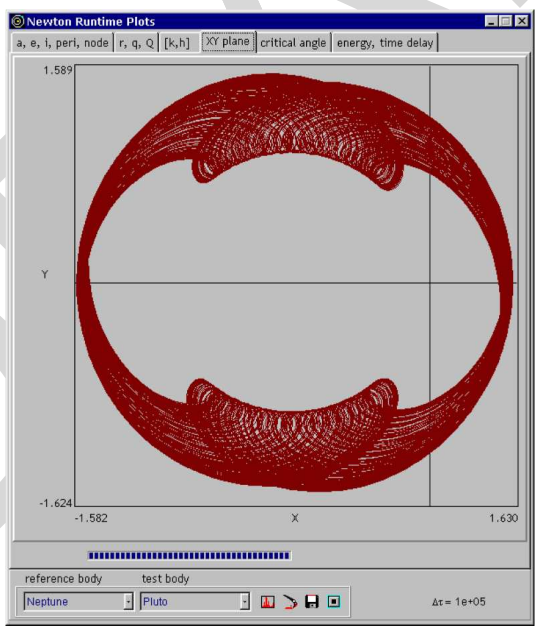


Figure 19: The orbit of Pluto over 10^5 years, as seen in a frame rotating with Neptune. (Neptune is located at the intersection of the crosshairs.) Pluto is in a 2:3 resonance with Neptune.

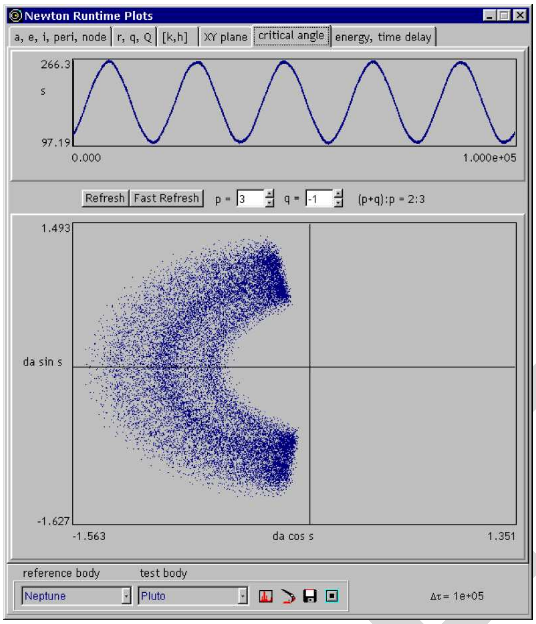


Figure 20: Critical angle for the Neptune-Pluto resonance.

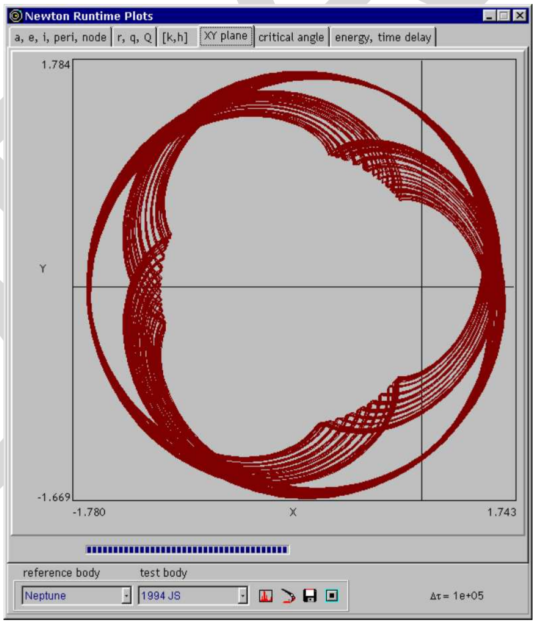


Figure 21: Orbit of the Trans-Neptunian Object 1994 JS, which is in a 3:5 resonance with Neptune.

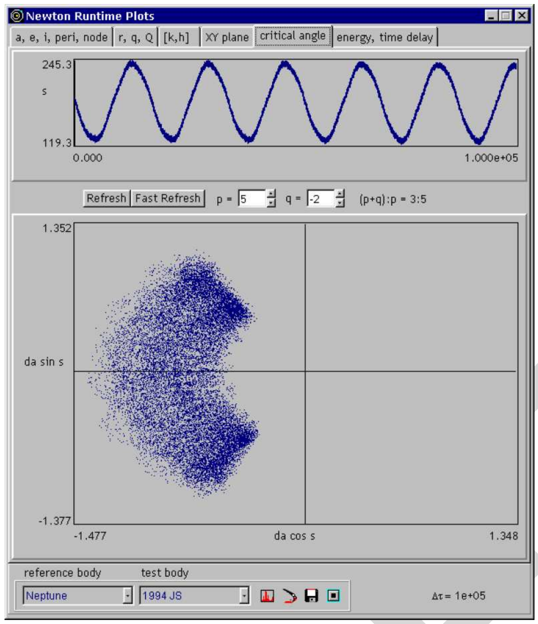


Figure 22: Critical angle for 1994 JS.

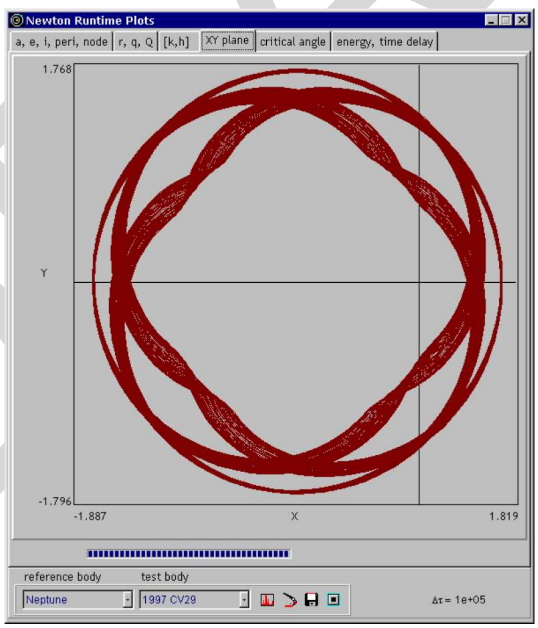


Figure 23: Orbit of 1997 CV_{29}, which is in a 4:7 resonance with Neptune.

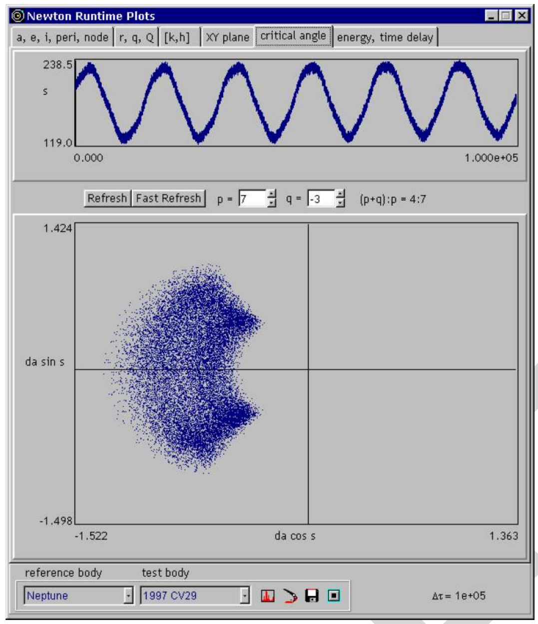


Figure 24: Critical angle for 1997 CV_{29}.

6.2.3 Resonances and Chaotic Asteroids

In this section, we make use of a certain diagnostic quantity that tells us something about the importance of chaos in the motion of a body. This diagnostic is called the Lyapunov exponent, λ. Suppose we have two particles that start off very near to each other, and we watch their subsequent dynamical evolution. In some cases, they will remain near each other for a very long time; their motion is, in some sense, stable. In other cases, the particles will separate, but only at a linear rate. After a certain time, their separation may be some amount Δr. After twice that amount of time, their separation will be something like $2\Delta r$. We call this *quasiperiodic* motion, and it, too, is in some sense stable. In some cases, however, the particles will separate at an exponential rate; after a relatively short time, they will be nowhere near each other. This is known as *sensitive dependence on initial conditions*, and this kind of motion we call *dynamical chaos*. This is from where the term, *butterfly effect*, which you've probably heard of, comes. We can characterize mathematically the separation Δr as a function of time as, say

$$\Delta r = f(t)e^{\lambda t} \tag{14}$$

where $f(t)$ is some function of time that is particular to the dynamical system being considered, and the term $e^{\lambda t}$ tells us if the motion is chaotic or not. If $\lambda > 0$ then we have chaos.

[...]

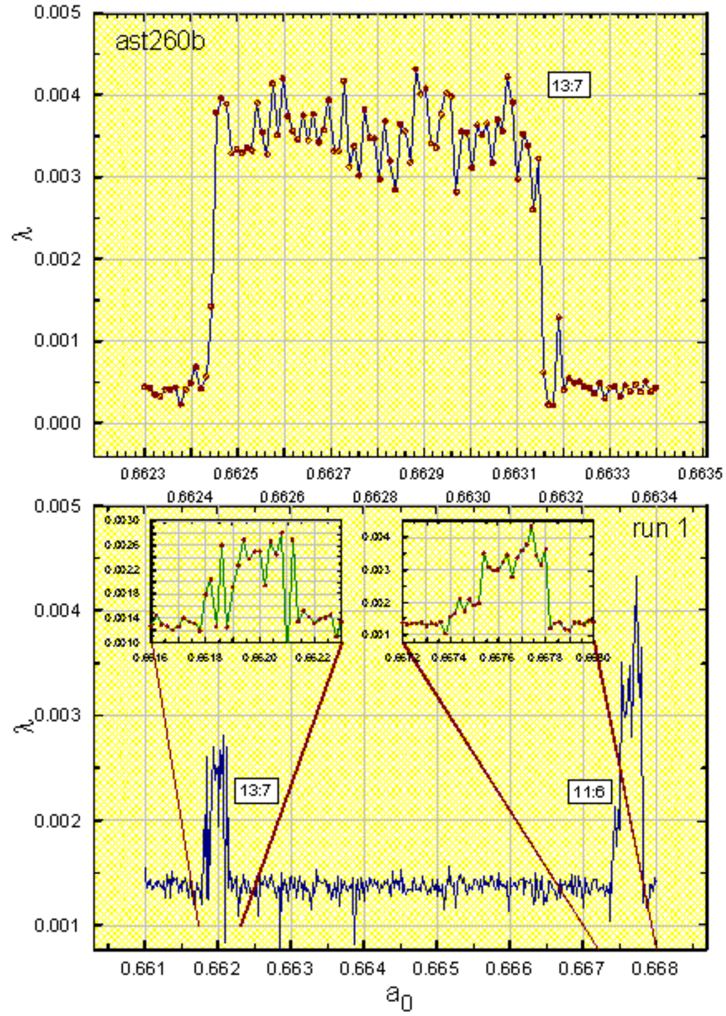


Figure 25: The largest Lyapunov exponent as a function of initial semimajor axis in a small section of the asteroid belt. Unit of distance is $a_{\mathrm{Jupiter}} = 5.203\,\mathrm{AU}$.

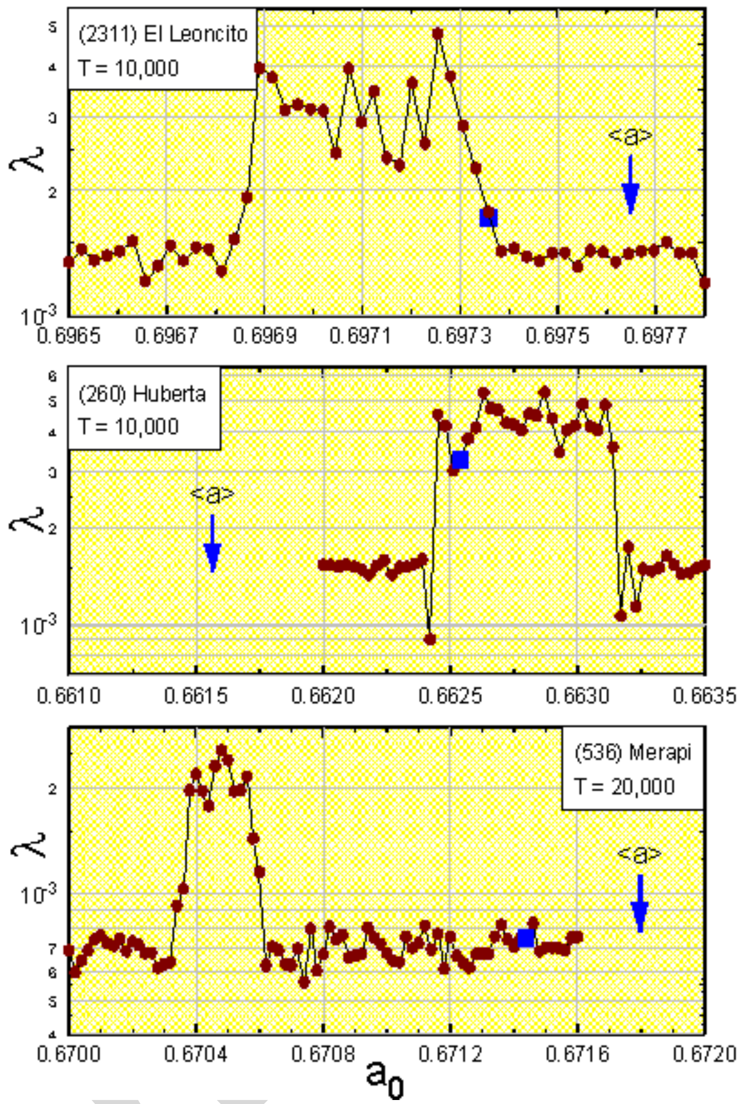


Figure 26: The largest Lyapunov exponent as a function of initial semimajor axis in vicinities of three asteroids. (2311) El Leoncito and (260) Huberta are in resonances. Unit of distance is $a_{\mathrm{Jupiter}} = 5.203$ AU.

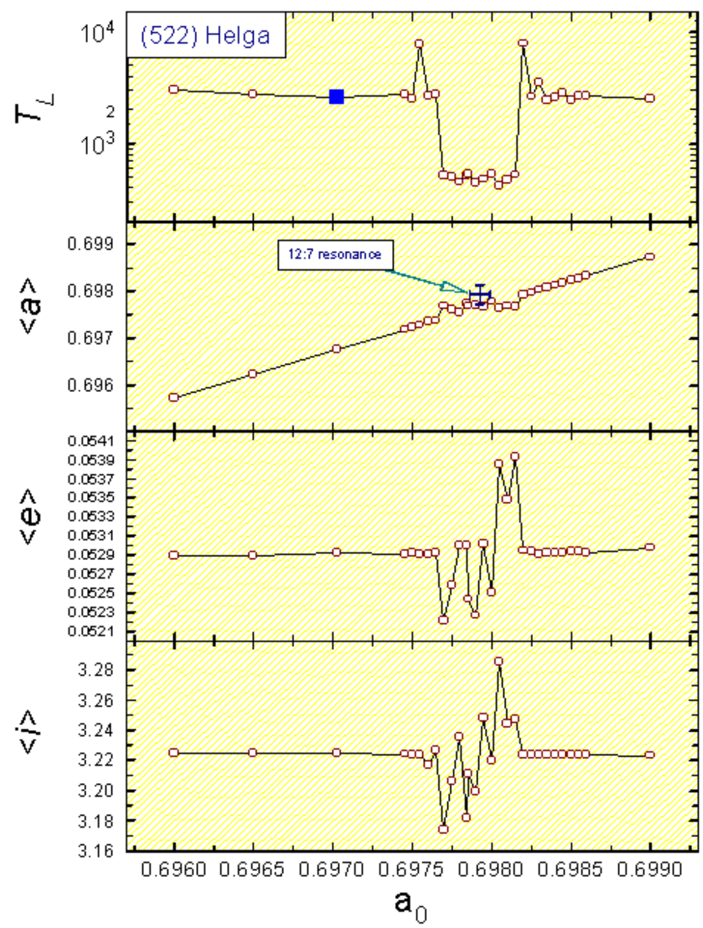


Figure 27: Mean orbital elements and the Lyapunov time $T_L = \frac{1}{\lambda}$ in the vicinity of the asteroid (522) Helga.

6.3 Secular Resonances

- definition
- delivery of asteroids to the inner solar system (NEOs)

6.4 Higher Order Mean Motion Resonances and the Origins of Chaos

- three-body MMRs
- origins of chaotic motion

7 Asteroid Noise

"*The stars make no noise.*" —Irish Saying

7.1 Two Kinds of Noise

Usually, when we think of noise, it has some unpleasant connotation. This is true in physics as it is in ordinary life. Noise invades every measurement – it is often very difficult to mitigate, and it is the ultimate barrier to the knowledge we seek from our measurements. Measurement noise is ordinarily most unwanted.

In addition to polluting measurements, noise is an intrinsic part of many physical processes. For example, you are probably familiar with the term *random walk*, which describes noise of a specific kind. The buffeting that a particular molecule receives from its neighbors in a gas or a liquid imparts just such a random walk to the molecule's motion. This is known as *Brownian motion*, and it has certain characteristics, such as the fact that the distribution of lengths of the short segments that make up its motion is Gaussian. There are other types of noise – Poisson noise, white noise, red noise, $1/f$ noise, etc. – with correspondingly different mathematical functions that characterize their distributions. Hence, we may turn the tables and, instead of cursing, use measurements of physical process noise to learn something about the system we're measuring. (We will still curse at measurement noise, alas.)

Taking a cue from the notion of process noise, we may adopt a somewhat unconventional view of planetary motions in the Solar System. Consider the simple, idealized two-body problem, a planet and the Sun. The planet's path in space is smooth and predictable. Now add another planet. Its presence will perturb the first planet's motion. Its path will still be smooth, but characterization and prediction in detail starts to become considerably more difficult. Continue adding perturbing masses, and the path in space, at small scales, becomes increasingly "wiggly" and unpredictable. The deviations from simple two-body motion begin to look more and more like process noise.

Now consider the planets of the inner solar system: Mercury through Mars. Their motions are perturbed by each other and by the outer planets, especially by massive (but, fortunately for us, distant) Jupiter. This web of perturbations certainly greatly complicates their individual paths through space. But not only are there the other planets, there are in addition hundreds of thousands of the small bodies we call asteroids. Now, the masses of individual asteroids are very small compared to the planets. However, their masses, individually and collectively, do affect planetary motions at length scales of order kilometers or less. Since there are many asteroids, the gravitational buffeting to which the inner solar system planets are subjected is, from a measurement perspective, indistinguishable from noise. Using Newton's equations of motion, we may explore and characterize the effects of "asteroid noise" on the motions of the planets. First, however, we should look at just what the distribution of the asteroids is, because that is the primary noise source, so to speak. We cannot understand the consequences of noise if we do not first understand the noise source itself.

7.2 The Asteroid Mass Distribution

Suppose we take the known asteroids and calculate masses for them, based on their observed brightness, their distance, and the fraction of sunlight that reflects from their surfaces. We can use this information to make a plot of the mass density as a function of position, such as that shown in Figures 28 and 29. The unit of distance is the astronomical unit, and the vertical scale in Fig. 28 is logarithmic. We immediately notice two features: the distribution is very lumpy, and the three largest asteroids, Ceres, Pallas, and Vesta, dominate over all others. Thus, as a crude approximation in calculating the orbits of the planets we might include just Ceres, or just the largest three, and ignore the rest. This would be a servicable model for many useful kinds of calculation.

In a more accurate consideration, however – one in which we are interested in the detailed effects of the asteroid population – we cannot ignore the lumpy nature of the asteroid mass distribution. If we remove all asteroids that are larger, say, than 200 km in diameter, we find the result shown in Figure 30. Since there are only 31 asteroids with $D > 200$ km, we might be tempted to integrate those 31 explicitly, along with the planets, and ignore the rest of the asteroids, hoping that this model will give a good approximation of the asteroid noise in the planetary motions. However, we see that the distribution in Figure 30 is also very uneven and lumpy. If we make the size cutoff 100 km, the leftover distribution is shown in Figure 31. We are starting to see some smoothing of the distribution, but it is still far from smooth. There are 264 asteroids with $D > 100$ km. Suppose we make the cutoff at 50 km in diameter. The result is Figure 32, where we see the distribution is finally becoming fairly smooth and even. In this case, we could then approximate the asteroids for which

$D < 50\,\text{km}$ as a smooth sheet of constant mass density, or even as a ring of mass at a distance equal to the density-weighted average and with a total mass equal to the sum of masses of these smaller bodies. Unfortunately, there are over 850 asteroids with $D > 50\,\text{km}$, and that is a large number to try and individually follow on a computer for any useful span of time (such as, for example, to study the noise in planetary motions).

In the next section (section 7.3, after the last of the mass density figures), we nevertheless discover that one can start to usefully characterize asteroid noise with much fewer than 850 individual asteroids.

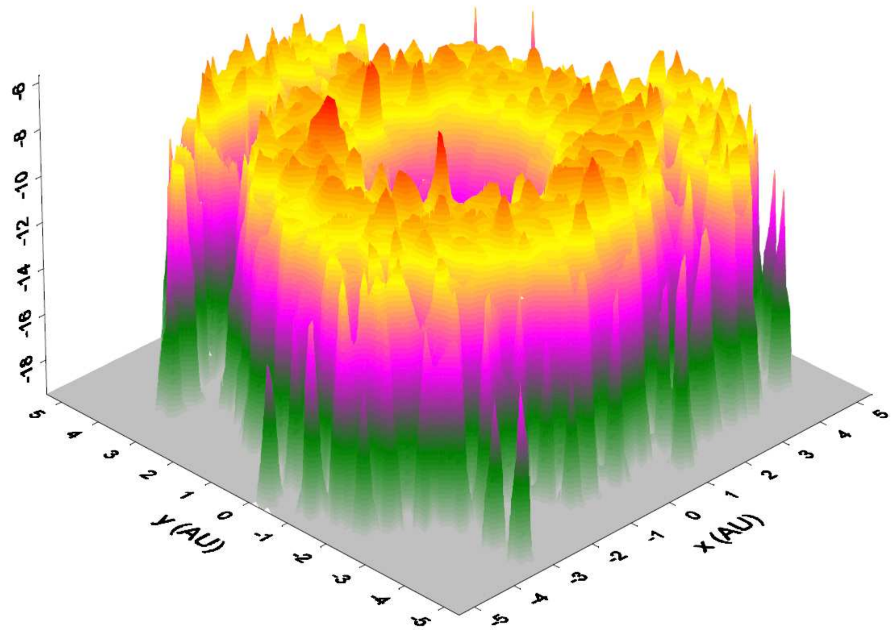


Figure 28: Mass density of the asteroids. Note that the scale is logarithmic. Ceres, Pallas, and Vesta stand out far above the rest of the asteroids.

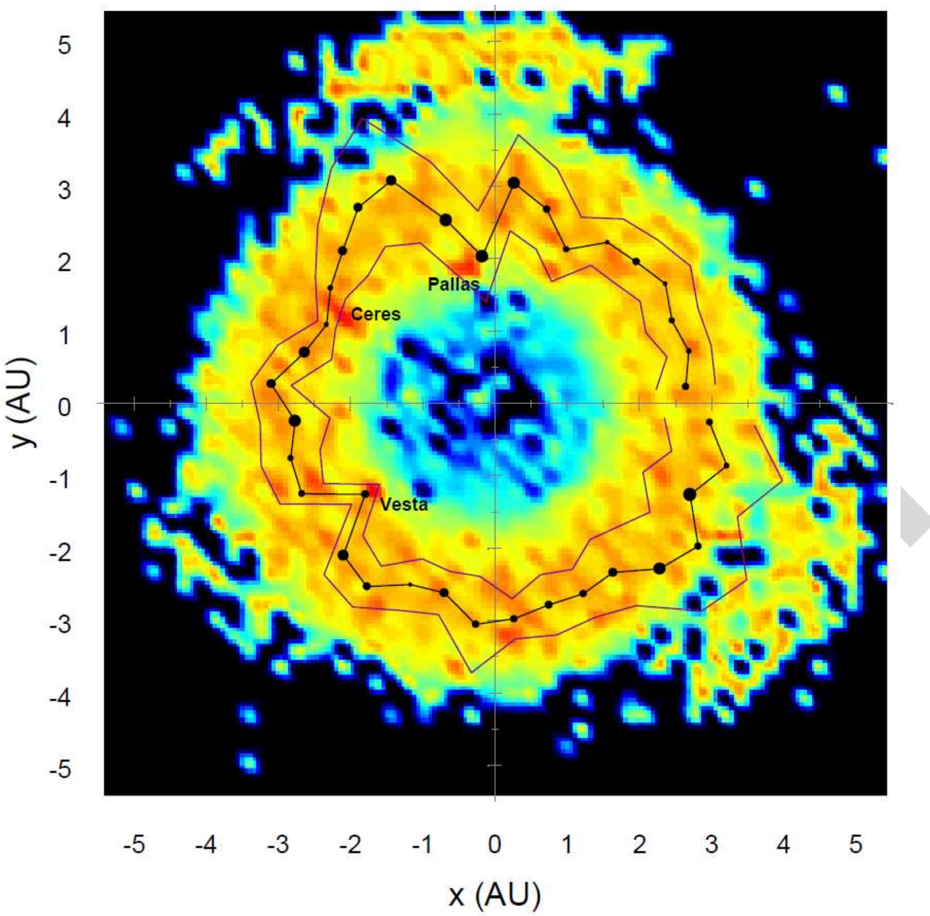


Figure 29: Two-dimensional plot of the asteroids mass density shown in Figure 28. The center line is the location of the radially-averaged mass density, $R = \frac{\iint r\rho(r,\theta)\ d\theta\, dr}{\iint \rho(r,\theta)\ d\theta\, dr}$, over a sector $\Delta\theta = \frac{\pi}{18}$. Note the Trojan asteroid clouds.

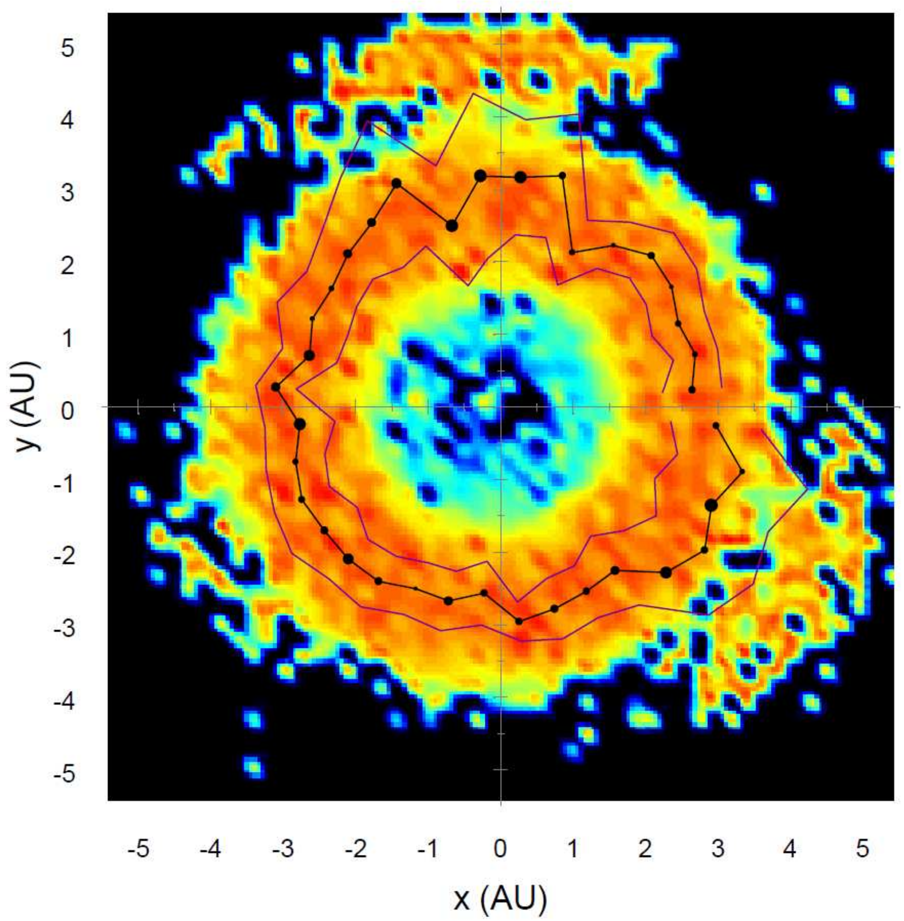


Figure 30: Same as Figure 29, but excluding all asteroids with $D > 200\,\mathrm{km}$.

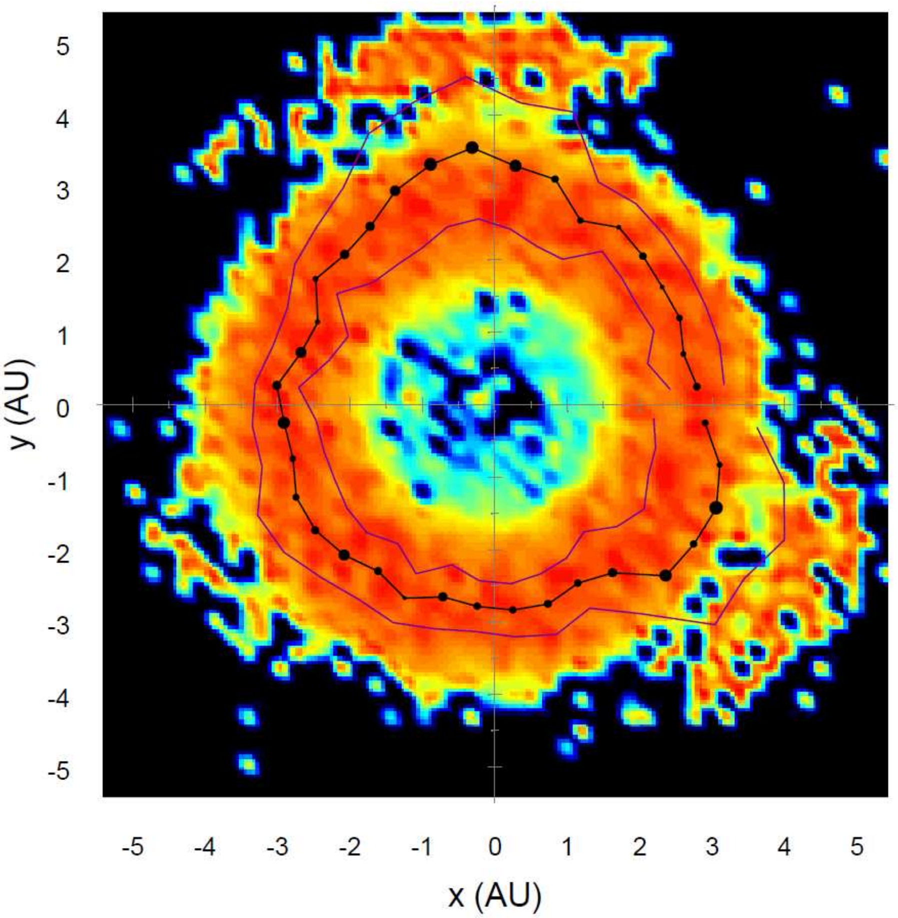


Figure 31: Same as Figure 29, but excluding all asteroids with $D > 100\,\text{km}$. Note the significant distortion in the radially-averaged position due to the Trojan clouds.

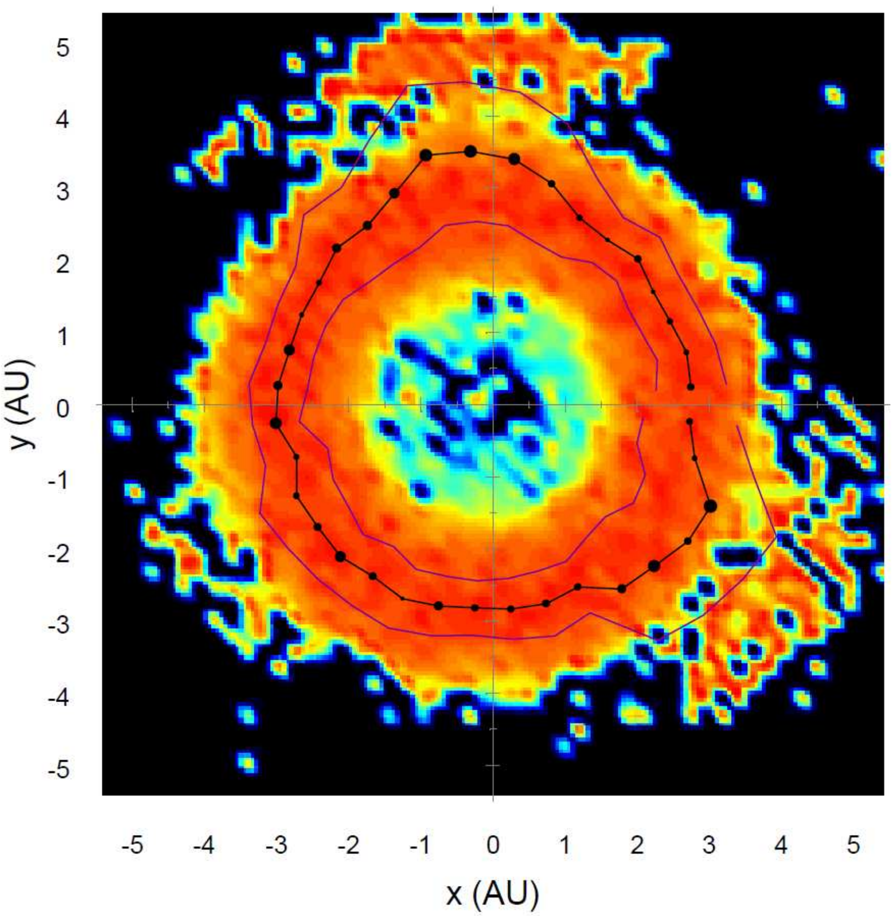


Figure 32: Same as Figure 29, but excluding all asteroids with $D > 50$ km. The distribution is finally becoming noticeably smoother.

7.3 The Noisy Motions of the Planets

7.3.1 Changes in the Orbital Elements

Let us now look at a few numerical integrations of the Newtonian equations of motion for the inner solar system planets, and see if we can deduce the effects of varying numbers of asteroids. In all cases for the results that follow, the positions and velocities of an integration of just the planets alone were subtracted from the positions and velocities of the integrations that included varying numbers of asteroids. Hence, the differences, or perturbations, are the quantities under consideration. From the positions and velocities, we may further state the results in terms deviations of the orbital elements due to the perturbing presence of asteroids. We can then study these deviations and learn something about asteroid noise.

Let's consider first perhaps the simplest case: that of Mars, the outermost inner Solar

System planet, being perturbed by Ceres, the most massive main-belt asteroid. [...]

In Figure 34, we look at the changes in semimajor axis of all the inner planets due to the "noise" from the 310 most-massive asteroids. [...]

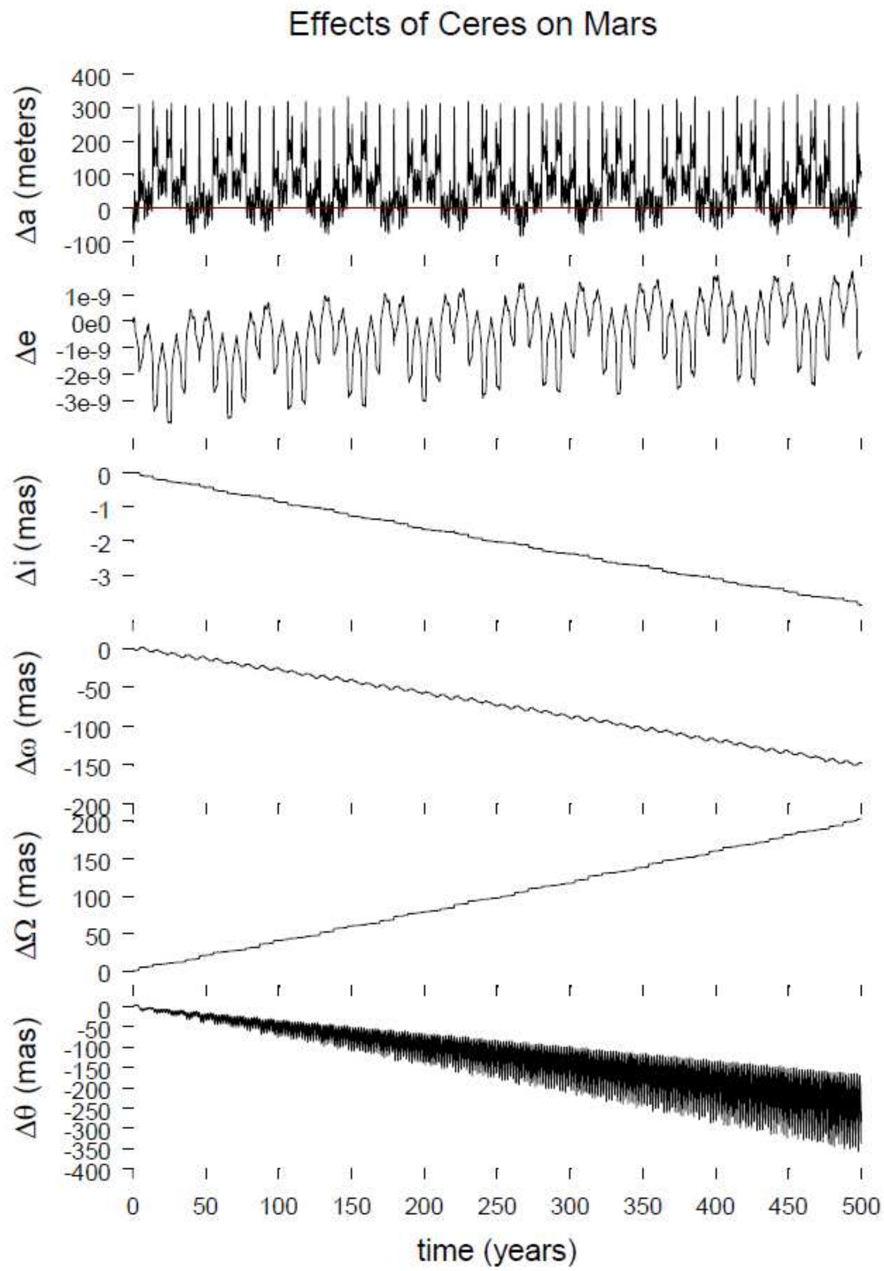


Figure 33: The effects of the asteroid Ceres on the orbital elements of Mars.

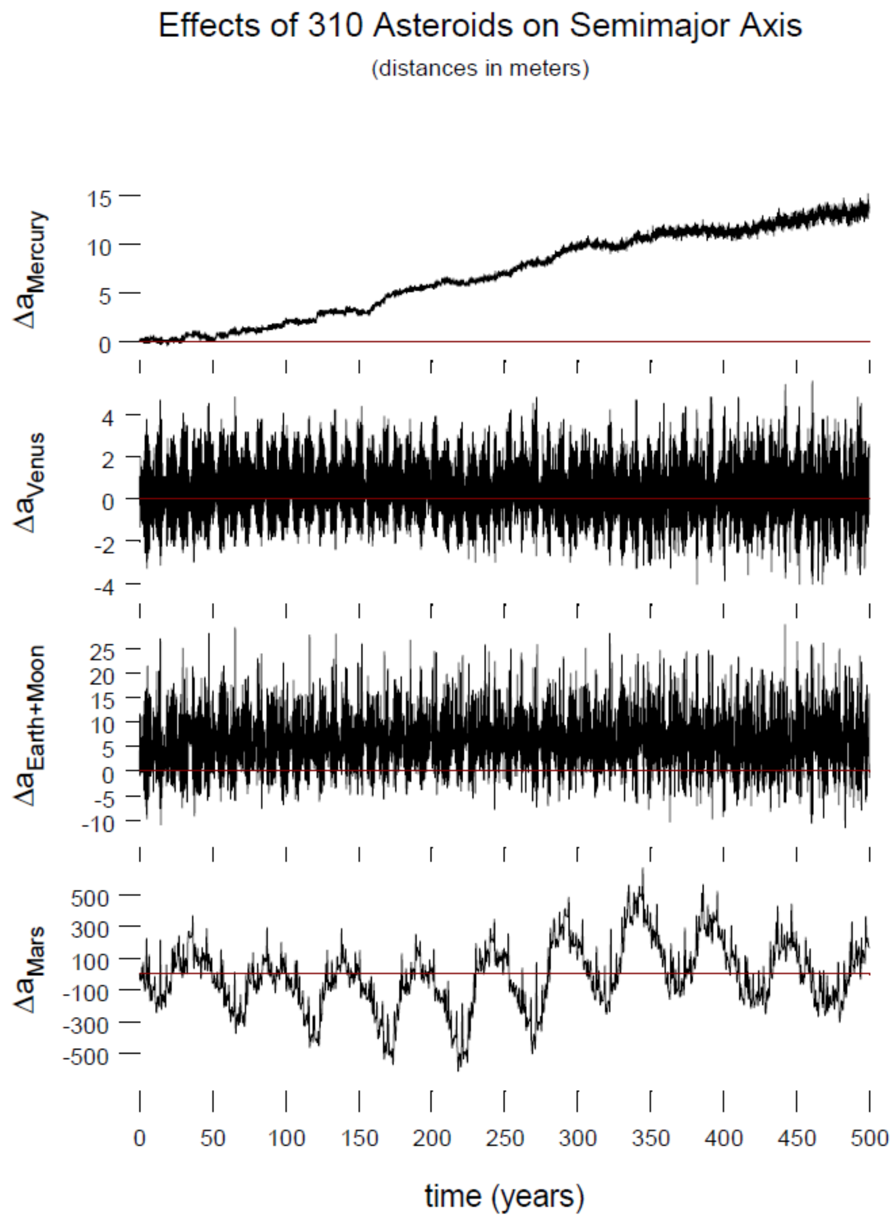


Figure 34: The effects of 310 asteroids ($D > 93\,\mathrm{km}$) on the semimajor axes of the inner Solar System planets.

7.3.2 Power Spectra

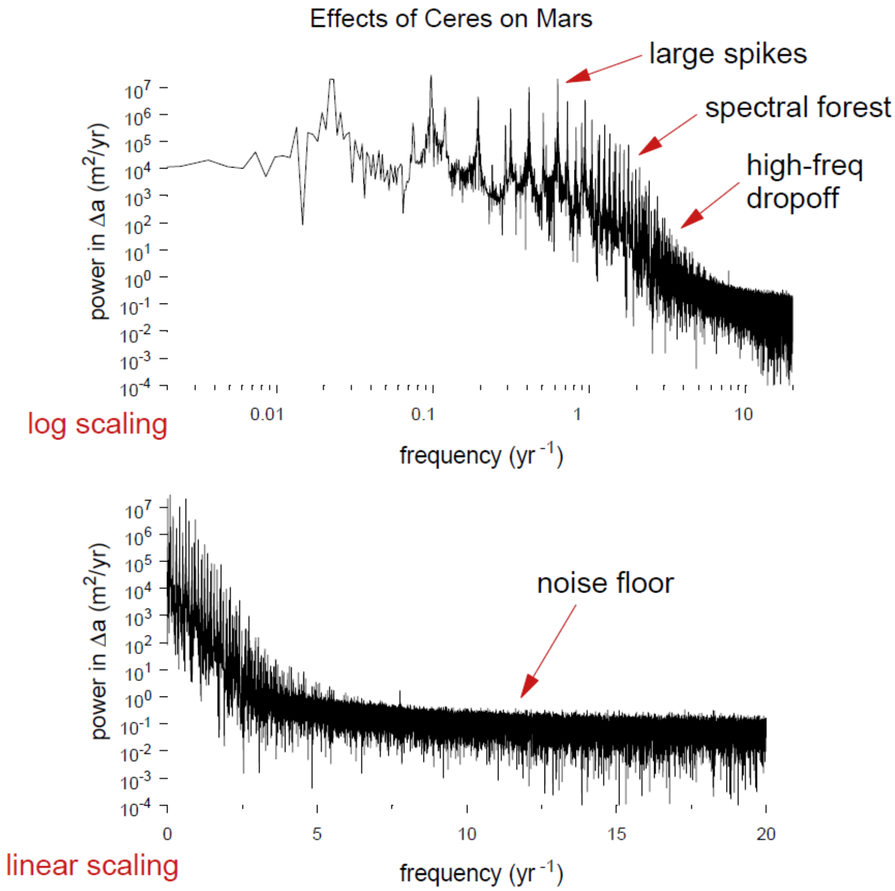


Figure 35: Power spectrum of the changes in semimajor axis of the orbit of Mars due to perturbations by Ceres.

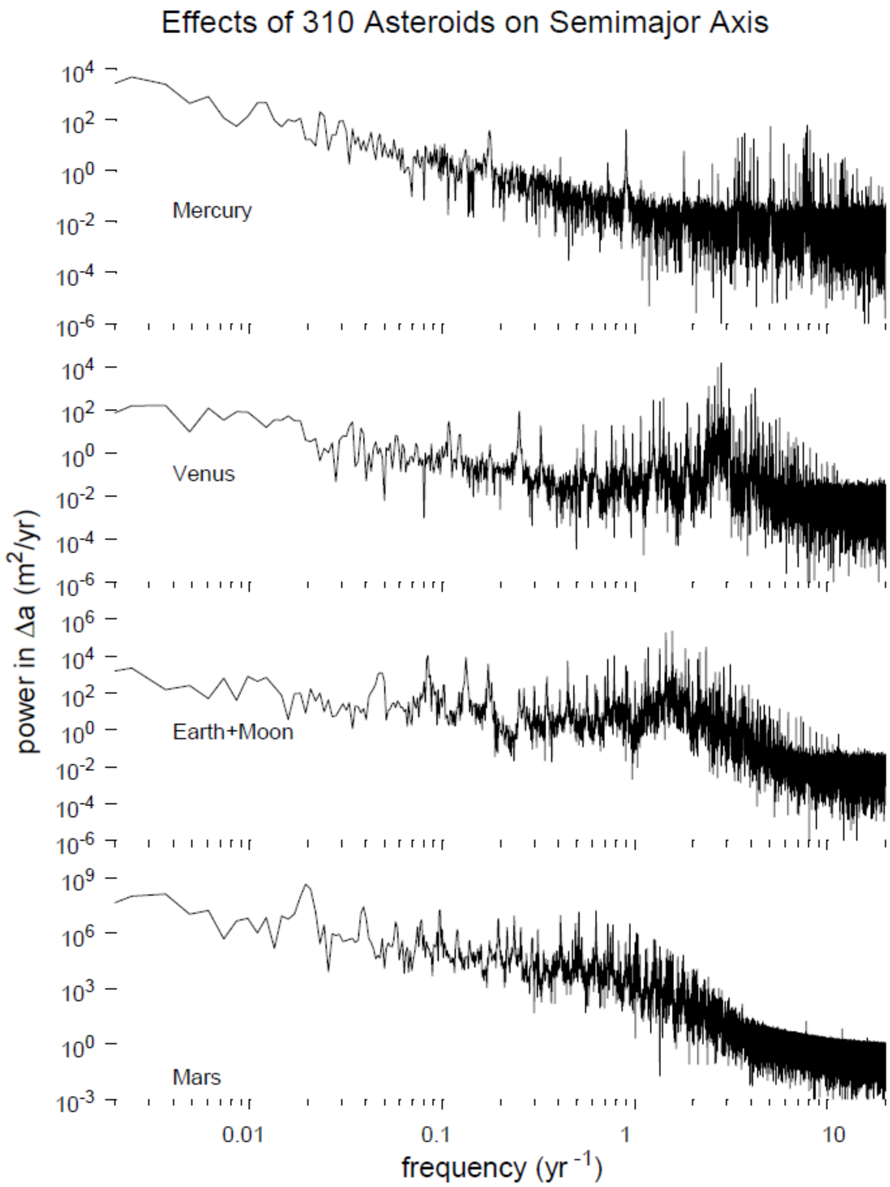


Figure 36: Power spectra of the changes in the semimajor axes of the inner planets due to perturbations by the 310 asteroids with diameter greater than 93 km.

7.3.3 The Destruction of Manifolds in Phase Space

[...]

Theorem 20. *(Takens embedding theorem) We can determine the geometric structure of the dynamics of a multivariate system from observations of a scalar diagnostic by constructing a time delay map from that scalar.*

[...]

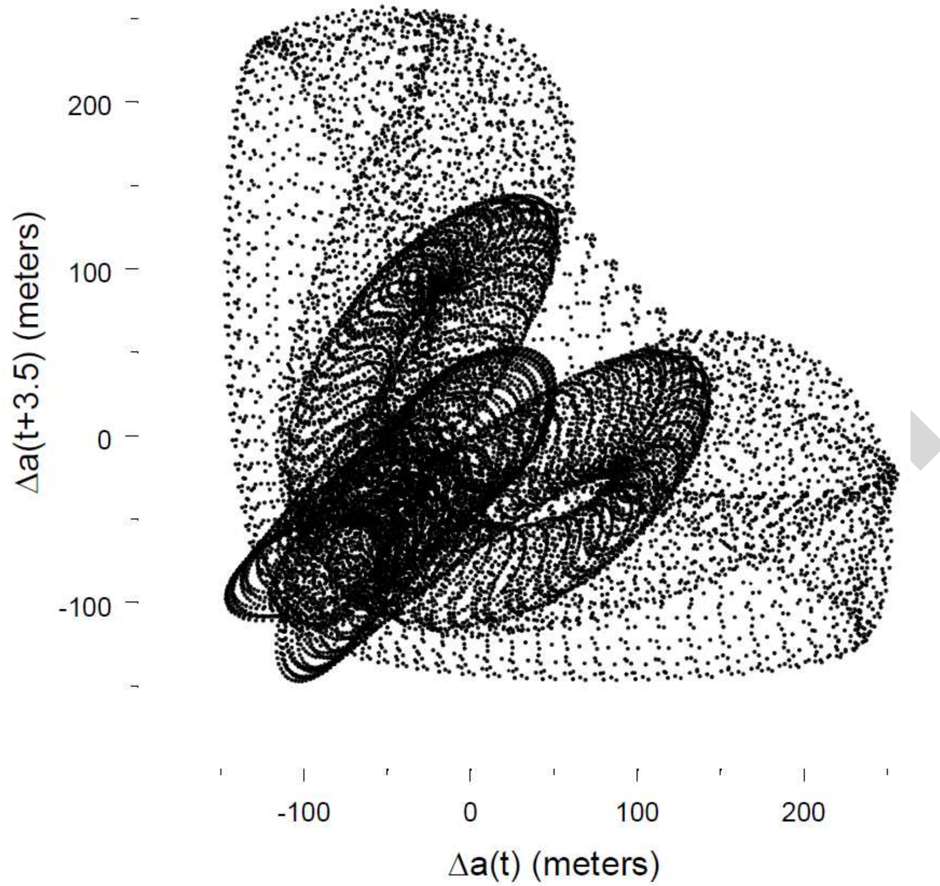


Figure 37: Phase space manifold on which Mars moves when perturbed by just Ceres (restricted three-body problem).

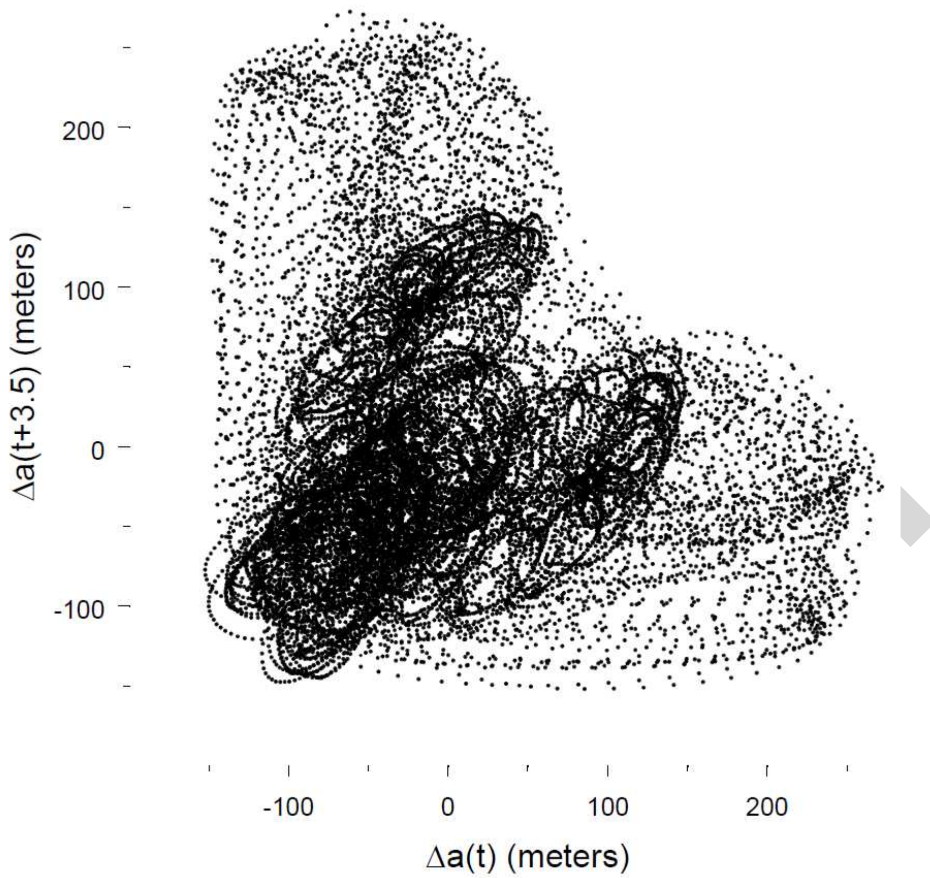


Figure 38: Same as Figure 37 but for Ceres plus the rest of the planets.

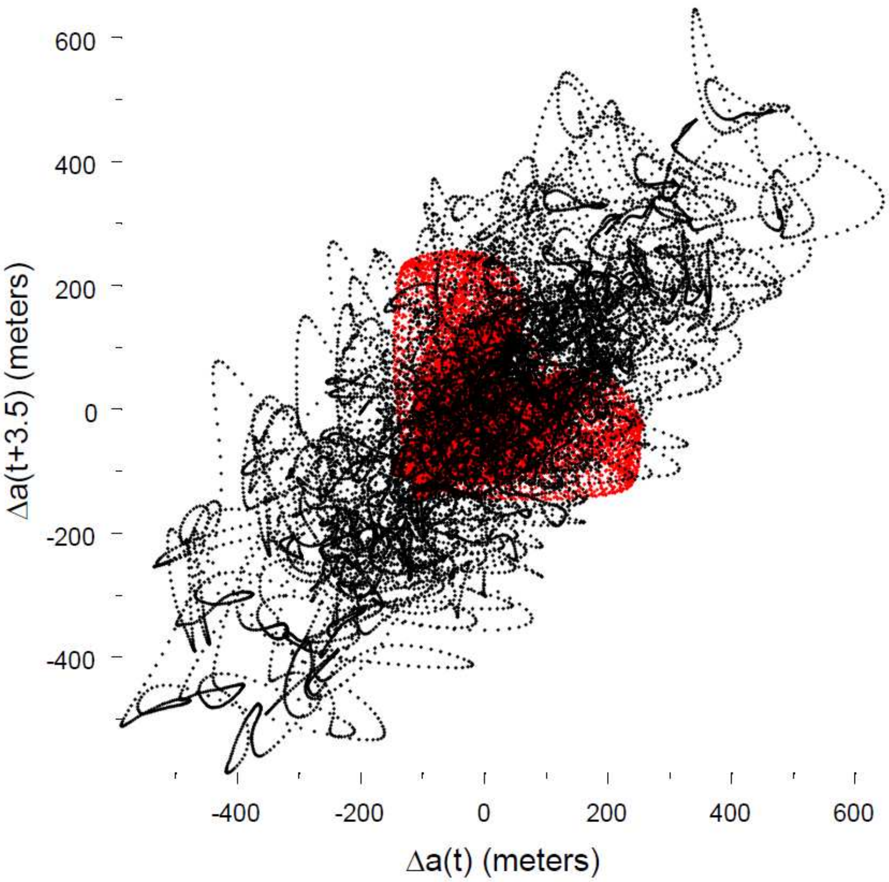
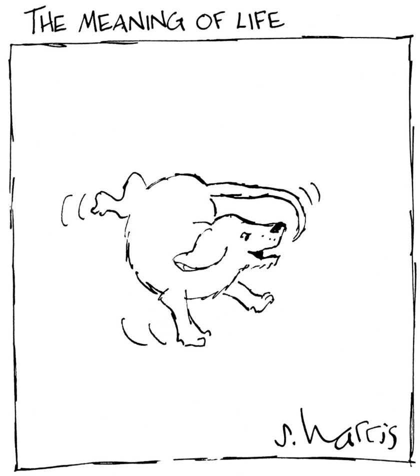


Figure 39: Same as Figure 37 but for the other planets plus all 310 of the largest asteroids ($D > 93$ km). The trace defined by the red dots is the structure from Figure 37.


THE MEANING OF LIFE
s. harris