The High Road/Low Road Demonstration or Birds on a Wire

Jacob Cady[1,*] and Chad A. Middleton[1,†]

[1]*Department of Physical and Environmental Sciences, Mesa State College, Grand Junction, CO 81501*

Abstract

Consider two separate tracks of equal horizontal displacements and equal initial and final heights. One track remains at this initial height while the other angles down, levels out, and then angles back up in order to regain its original height. Question: If two identical balls are set rolling with equal initial speeds, which ball completes the track in a shorter time interval? In this manuscript, the dynamics of a ball on each track are analyzed using basic Newtonian mechanics. We calculate the time necessary to complete each path in terms of the parameters of the track and the initial velocities of the balls. We derive an expression for the time difference between the two tracks and compare this to data taken on a set of high road/load road tracks, hence demonstrating the fact that the ball traversing the low road *always* wins the race.

Key words: conservation of energy, kinematics, moment of inertia, Newtonian mechanics, rolling without slipping

*Electronic address: jlcady@mesastate.edu
†Electronic address: chmiddle@mesastate.edu

I. INTRODUCTION

Consider two separate tracks of equal horizontal distance and initial and final heights. The flat track remains at this constant initial height while the second track dips down into a flat valley, hence following a longer path, before regaining its original height. An example of such a set of tracks is displayed in Fig. 1. Question: If two identical balls are set rolling with equal initial speeds, which ball completes the track in a shorter time interval? The ball that takes the deviated path into the valley will of course travel at a higher speed for much of the race, however, that ball has a longer distance to travel. On first glance, the answer may not seem at all obvious. One of us first encountered the aforementioned question/ demonstration at the *13th Workshop for New Physics and Astronomy Faculty* where the workshop attendees failed to reach a consensus on the correct outcome prior to witnessing the demonstration. In fact, Leonard et. al. found that about two-thirds of introductory physics students wrongly predict that both balls will complete the track in equal times. One reason often cited for this prediction involves invoking conservation of energy to (correctly) identify that the ball will have the same initial and final speeds and conclude (incorrectly) that the balls must therefore tie. Further, about one-sixth of intro students incorrectly predict that the ball following the flat track will reach the other end first (as they travel the shorter distance). Only about one-tenth correctly identify that the ball following the longer path will in fact win the race [1]. A further study of students' judgments concerning the outcomes of races involving such a pair of tracks is presented in [2].

The answer to the above conceptual question becomes obvious when the velocity components are analyzed. When the dissipative effects of friction are ignored, the speed of the ball traversing the flat track remains constant. For the ball that follows the longer path, the horizontal component of the velocity is always equal to or larger than that of the speed of the ball on the flat track. This is due to the fact that the ball experiences a horizontal net force on the first diagonal section of the track causing the ball to undergo a horizontal acceleration. Although the ball does in fact experience a horizontal deceleration on second diagonal section, the ball's horizontal component of velocity will always remain at or above the initial speed of the ball. *Thus, the ball traversing the longer path always wins the race, regardless of the initial speed.* This outcome is also independent of the breadth of the flat valley with a similar result emerging from a V-valley track [1, 2].

As is well known, Johann Bernoulli (1667-1748) found the solution of the path of shortest travel time between two points for an object under the influence of a constant gravitational force. This pioneering work began the study of the calculus of variations. In fact, the brachistochrone is often first introduced in an advanced undergraduate dynamics class as an example of an application of the calculus of variations. Since this original work, there have been several follow up analyses of generalizations of the brachistochrone and tautochrone including [3–11]. Several articles have been written highlighting the incorporation of the brachistochrone into the classroom, see for example [12–14].

Zheng et. al. considered a race involving beads of equal initial speeds on a straight and curvilinear path [15] and found that the outcome of the race was dependent on the initial speed (see [14] for a similar result involving spherical balls on a straight path vs. the brachistochrone curve). This result differs from that of the high road/low road where the ball on the low road always wins, independent of the initial speed. This apparent inconsistency arises from the fact that the scenario of Zheng et. al. is fundamentally different than that of the high road/low road and can be understood in terms of the orientation of the initial velocity

vectors. For two non-parallel initial velocity vectors of equal magnitudes, the horizontal components differ and thus allow for a race dependent on the initial speed.

In section II, we calculate the time interval for a spherical ball to traverse each track in terms of the parameters of the track and the initial speed of the ball. From this, one can easily construct the time difference. The derivation of the time interval of the low road consists of dissecting the track into five distinct regions and calculating the time needed to traverse each. In section III, we discuss the experimental design and setup of our high road/low road apparatus. In section IV, we discuss our experimental procedure and present the results as a plot of the time difference versus the average speed of the ball.

FIG. 1: An example of a high road/low road track

II. THEORETICAL ANALYSIS

The two tracks studied here are equal in horizontal displacement and are symmetrical about their midpoints. Both tracks begin and end at the same initial height ($h_{high,i} = h_{low,i} = h_{high,f} = h_{low,f}$). The tracks differ in that one dips down in its midsection while the height of the other remains constant. Throughout this paper, *low road* will refer to the track that does in fact deviate in the midsection; *high road* will refer to the track that remains at constant height. A diagram of the first half of the low road is illustrated as the red line in Fig. 2, which includes other relevant track parameters.

Each symmetric half of the low road is partitioned into five sections. The upper horizontal is the first flat reach that is level with the high road. This track then enters an area of constant radius of curvature, where the track deviates from the upper horizontal, before beginning the straight diagonal section. The track then enters another area of constant radius of curvature before reaching the second flat reach of the lower horizontal section. The radii of curvature of both arcs are identical and were conveniently designed and constructed to be equal to the elevation difference between the higher and lower horizontal sections of the low road.

A. The Horizontal Sections of the High and Low Road

We begin by first calculating the time needed to traverse the high road. As the ball undergoes zero acceleration, the time is simply given by

$$t_H = \frac{L}{v_1}, \tag{2.1}$$

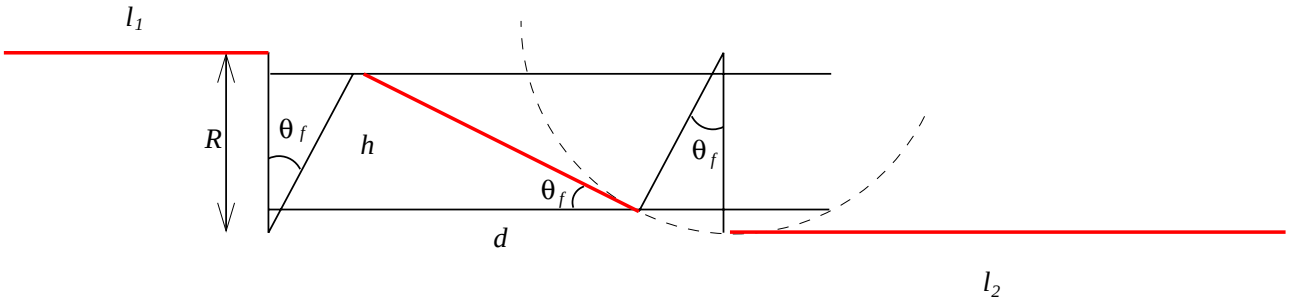


FIG. 2: Schematic of the first half of the low road.

where v_1 is the initial velocity of the ball. L is the total length of the track and can be written in terms of the parameters of the low road as

$$L = 2(l_1 + l_2 + d + 2R\sin\theta_f). \tag{2.2}$$

Calculating the time needed for the ball to traverse the low road is inherently more involved. As the track is symmetrical and the dissipative effects of rotational friction will be ignored in this theoretical treatment, we can calculate the time necessary to traverse half of the low road and multiply by a factor of two to obtain the total time. Furthermore, we divide the first half of the track into its five constituent parts, as can be seen in Fig. 2, and analyze each part in turn. The total time needed to traverse the low road, consequently, will be the sum of these individual times given by the relation

$$t_L = 2\sum_{i=1}^{5} t_i. \tag{2.3}$$

As was the case for the entire high road, the time necessary for the ball to traverse the upper horizontal section of the low road is the length of the segment divided by the velocity. This time is given by

$$t_1 = \frac{l_1}{v_1}. \tag{2.4}$$

We next analyze the lower horizontal section. In order to calculate the time needed for the ball to traverse the lower horizontal, we must first find the speed of the ball on this section. We find this speed by using the fact that the total mechanical energy is conserved, where again any energy loss due to rotational friction is neglected. Applying conservation of mechanical energy yields the expression

$$mgR + \frac{1}{2}mv_1^2 + \frac{1}{2}I\omega_1^2 = \frac{1}{2}mv_2^2 + \frac{1}{2}I\omega_2^2, \tag{2.5}$$

where the moment of inertia of a spherical ball is

$$I = \frac{2}{5}mR_b^2, \tag{2.6}$$

with R_b the proper radius of the ball. The angular velocity of the ball is proportional to the translational velocity and is

$$\omega = \frac{v}{R_e}, \tag{2.7}$$

where R_e is the effective radius of the ball, which differs from the proper radius. As the ball lies on two rails, the radius that connects the translational velocity to the angular velocity is the vertical distance between the point where the ball touches the rail and the center of the ball. See Appendix A for a more detailed discussion and derivation of this effective radius in terms of the proper radius.

Inserting Eqs. (2.6) and (2.7) into Eq. (2.5), we obtain the expression

$$mgR + \frac{1}{2}m\left[1 + \frac{2}{5}\left(\frac{R_b}{R_e}\right)^2\right]v_1^2 = \frac{1}{2}m\left[1 + \frac{2}{5}\left(\frac{R_b}{R_e}\right)^2\right]v_2^2. \tag{2.8}$$

Now, solving for the velocity of the ball on the lower horizontal in terms of the initial velocity is given by

$$v_2 = \sqrt{v_1^2 + 2g'R}\,, \tag{2.9}$$

where

$$g' \equiv g\left[1 + \frac{2}{5}\left(\frac{R_b}{R_e}\right)^2\right]^{-1} \tag{2.10}$$

plays the role of an effective gravitational acceleration and will be used throughout this paper. Having an expression for the velocity of the ball on the lower horizontal allows for a calculation of the time needed to traverse the lower horizontal in terms of the initial velociy. This time, in terms of the initial velocity of the ball and the parameters of the track, is given by the relation

$$t_2 = \frac{l_2}{\sqrt{v_1^2 + 2g'R}}. \tag{2.11}$$

B. The Upper Constant Radius of Curvature Section of the Low Road

The tangential acceleration of the center of mass of the ball on the upper constant radius of curvature section is dependent upon the angle subtended along the arc. Fig. 3 illustrates this scenario and includes a free-body diagram of the forces acting on the ball.

Newton's 2nd Law for translational motion in the centripetal and tangential directions are, respectively

$$mg\cos\theta - n = \frac{mv^2}{r} \tag{2.12}$$

$$mg\sin\theta - f_r = ma_{cm}, \tag{2.13}$$

where $r = R + R_e$ is the distance from the center of the radius of curvature of the track to the center of the spherical mass. f_r is the static friction acting by the rails on the ball that produces rotation, a_{cm} is the acceleration of the center of mass of the ball. Notice that our analysis assumes that the ball never leaves the track. This remains true so long as the speed of ball is less than $\sqrt{gr\cos\theta}$. For speeds larger than this, the normal force on the ball becomes zero and the ball leaves the track.

The acceleration of the center of mass is related to the angular quantities by

$$a_{cm} = r\ddot{\theta} = R_e\ddot{\phi}, \tag{2.14}$$

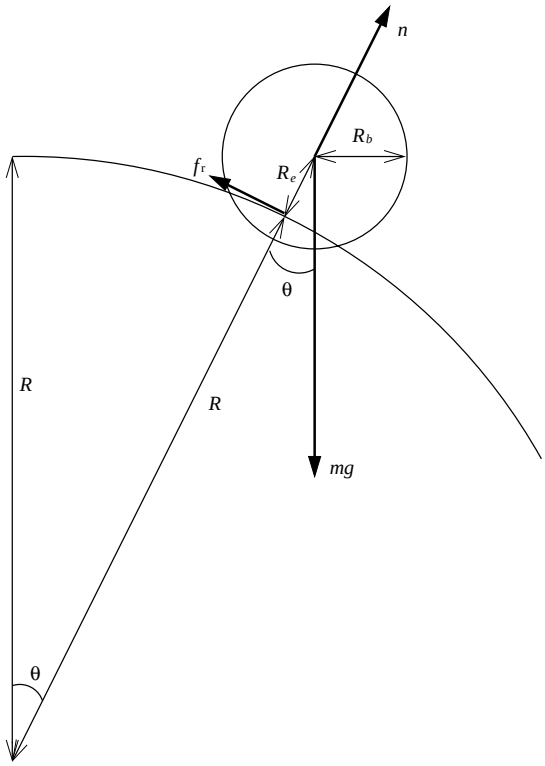


FIG. 3: Schematic of the ball traversing the upper constant radius of curvature section.

where $\ddot{\phi}$ is the angular acceleration of the ball about its center of mass. Newton's 2nd Law for rotational motion yields the expression

$$f_r R_e = \frac{2}{5} m R_b^2 \ddot{\phi}. \tag{2.15}$$

Substituting Eqs. (2.15) and (2.14) into Eq. (2.13) and rearranging, we arrive at the dynamical expression

$$\ddot{\theta} = \dot{\theta}\frac{d\dot{\theta}}{d\theta} = \frac{g'}{r}\sin\theta, \tag{2.16}$$

where we used the chain rule to connect the angular acceleration of the ball to its angular velocity. One can now separate variables and integrate to arrive at an expression for the angular velocity

$$\dot{\theta}^2(\theta) = \frac{v_1^2}{r^2} + \frac{2g'}{r}\left(1 - \cos\theta\right), \tag{2.17}$$

where we applied the boundary condition $\dot{\theta}(0) = v_1/r$ to obtain a value for the integration constant. The final velocity of the ball on the constant radius of curvature section takes the form

$$v(\theta_f) = \sqrt{v_1^2 + 2g'r(1 - \cos\theta_f)}, \tag{2.18}$$

which, consequently, will be the ball's initial speed on the diagonal section of the track.

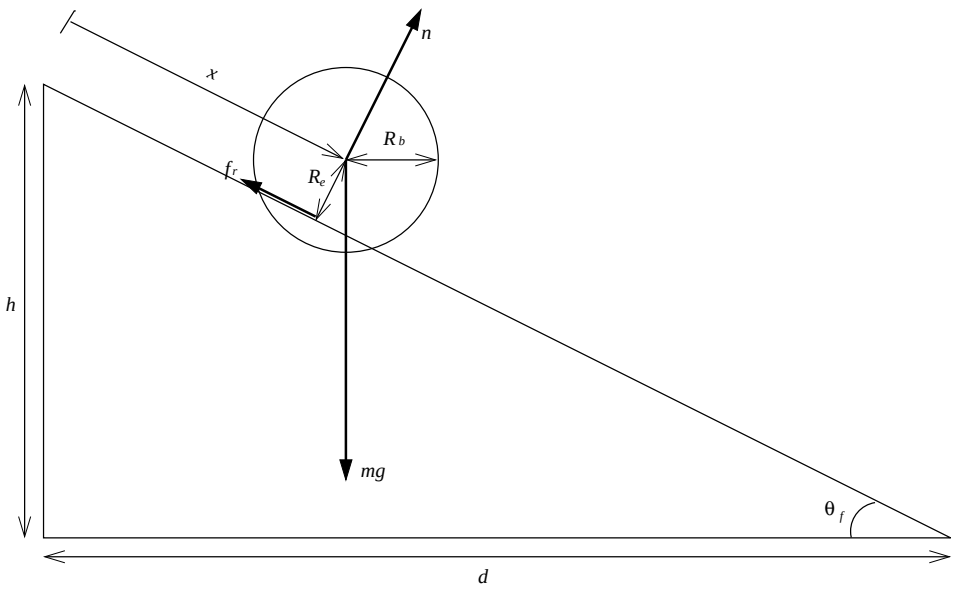


FIG. 4: Schematic of the ball traversing the diagonal section of the track.

Eq. (2.17) can be further separated and integrated. One arrives at an integral expression for the time spent on the constant radius of curvature section given by

$$t = \int_0^{\theta_f} \frac{r\ d\theta}{\sqrt{v_1^2 + 2g'r(1-\cos\theta)}}. \tag{2.19}$$

This integral is exactly solvable with the solution a hypergeometric function. To keep this analysis as simplistic as possible, we instead employ a small angle approximation and solve the integral perturbatively. To lowest order in θ

$$1 - \cos\theta \simeq \theta^2/2. \tag{2.20}$$

For our track, this approximation amounts to a maximum percent error, when $\theta = \theta_f$, of $\sim$1%. Using the above approximation, the integral yields an expression for the time spent on the constant radius of curvature section of the track

$$t_3 \simeq \sqrt{\frac{r}{g'}} \sinh^{-1} \quad \frac{\sqrt{g'r}}{v_1}\ \theta_f \quad . \tag{2.21}$$

C. The Diagonal Section of the Low Road

Applying Newton's 2nd Law to the ball on the diagonal in the parallel/perpendicular directions yields

$$n - mg\cos\theta_f = 0 \tag{2.22}$$
$$mg\sin\theta_f - f_r = m\ddot{x}, \tag{2.23}$$

where x is the location of the center of mass of the ball along the diagonal path, as shown in Fig. 4. Assuming that the ball does not slip on the diagonal, the coordinates x and ϕ are

proportional and related by

$$\phi = \frac{x}{R_e}, \tag{2.24}$$

where ϕ is the angular displacement of the ball.

Newton's 2nd Law for rotational motion is again given by

$$f_r R_e = \frac{2}{5} m R_b^2 \ddot{\phi}. \tag{2.25}$$

Substituting Eqs. (2.24) and (2.25) into Eq. (2.23) and rearranging yields the dynamical expression

$$\ddot{x} = g' \sin\theta_f, \tag{2.26}$$

where the effective acceleration of gravity, g', was defined in Eq. (2.10).

As the acceleration of the ball on the incline is constant, we can invoke the kinematic equations

$$\begin{aligned} v &= v_i + a\Delta t \\ \Delta x &= v_i \Delta t + \frac{1}{2} a \Delta t^2 \\ v^2 &= v_i^2 + 2a\Delta x \end{aligned} \tag{2.27}$$

to find the time needed to traverse the diagonal section. The initial velocity, v_i, of the ball on the diagonal is known in terms of the initial velocity, v_1, of the ball at the beginning of the track. Eq. (2.18) yields

$$v_i = \sqrt{v_1^2 + 2g'r(1-\cos\theta_f)}. \tag{2.28}$$

Plugging this into Eq. (2.27) gives the speed of the ball at the end of the diagonal section

$$v(x = \sqrt{h^2+d^2}) = \sqrt{v_1^2 + 2g'[r(1-\cos\theta_f)+h]}. \tag{2.29}$$

We now obtain an expression for the time to traverse the diagonal section of the track given by

$$t_4 = \frac{1}{g'\sin\theta_f}\left[\sqrt{v_1^2 + 2g'[r(1-\cos\theta_f)+h]} - \sqrt{v_1^2 + 2g'r(1-\cos\theta_f)}\,\right] \tag{2.30}$$

in terms of the parameters of the track and the initial velocity of the ball.

D. The Lower Constant Radius of Curvature Section of the Low Road

A derivation of the time for the ball to traverse the lower constant radius of curvature section closely parallels that of the upper constant radius of curvature section. Newton's 2nd law for translational motion in the centripetal and tangential direction are

$$n - mg\cos\theta = \frac{mv^2}{\tilde{r}} \tag{2.31}$$

$$mg\sin\theta - f_r = ma_{cm} \tag{2.32}$$

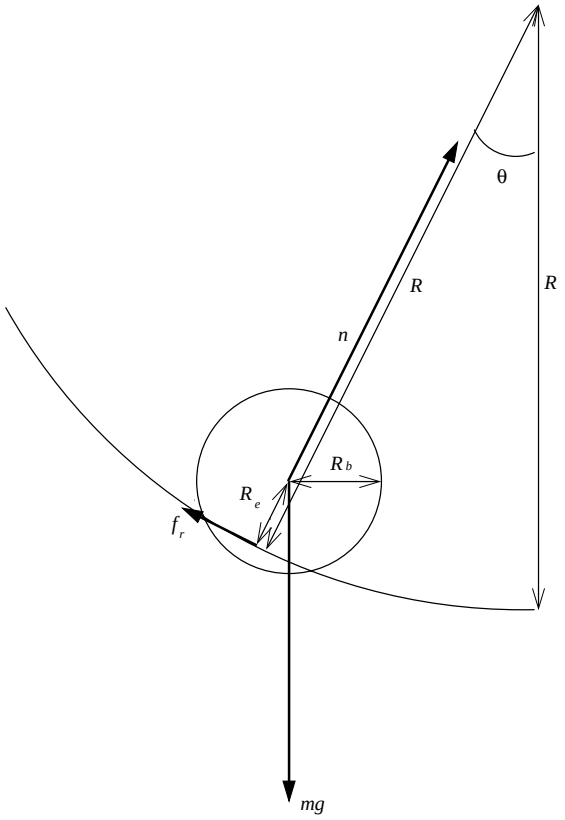


FIG. 5: Schematic of the ball traversing the lower constant radius of curvature section.

where $\tilde{r} = R - R_e$ is the distance from the center of the radius of curvature of the track to the center of the spherical mass. Notice that θ is decreasing as ϕ, the angular displacement of the ball about its center of mass, is increasing. Thus, our angular quantities are related by where

$$a_{cm} = R_e\ddot{\phi} = -\tilde{r}\ddot{\theta}. \tag{2.33}$$

Newton's 2nd law for rotational motion is of the form

$$f_r R_e = \frac{2}{5} m R_b^2 \ddot{\phi} \tag{2.34}$$

Inserting Eq. (2.34) for f_r into Eq. (2.32) and rearranging, one obtains an expression of the form

$$\ddot{\theta} = -\frac{g'}{\tilde{r}} \sin\theta \tag{2.35}$$

Now following a process similar to that of section IIB, we find the time needed to traverse the lower constant radius of curvature section is

$$t_5 \simeq \sqrt{\frac{\tilde{r}}{g'}}\ \sin^{-1}\left(\sqrt{\frac{g'\tilde{r}}{v_1^2 + 2g'R}}\ \theta_f\right), \tag{2.36}$$

where we set our limits of integration from $-\theta_f$ to 0, again used a small angle approximation, and applied the final boundary condition Eq. (2.9) for the velocity. The velocity expressed in Eq. (2.9) is that of the ball's as it traverses the lower horizontal section of the track.

E. Total Time Difference vs. Initial Speed

We can now calculate the total time difference between transits of the high and low road for a given initial velocity. The total time difference is defined as the time interval needed for the ball to traverse the high road minus the time interval needed for the ball to traverse the low road

$$\Delta t = t_H - t_L. \tag{2.37}$$

As the ball traversing the low road *always* wins the race, one expects Eq. (2.37) to be positive for all values of the initial velocity. In a previous subsection, it was found that the time for the ball to traverse the high road was given by

$$t_H = \frac{2(l_1 + l_2 + d + 2R\sin\theta_f)}{v_1}. \tag{2.38}$$

The time needed for the ball to traverse the low road is found by summing the time contributions from each of the five sections of the track and multiplying by a factor of two. The expression is of the form

$$\begin{aligned} t_L \;=\; 2 \;\; & \frac{l_1}{v_1} + \frac{l_2}{\sqrt{v_1^2 + 2g'R}} + \;\; \sqrt{\frac{r}{g'}}\sinh^{-1} \;\; \frac{\sqrt{g'r}}{v_1}\,\theta_f \;\; + \;\; \sqrt{\frac{\tilde{r}}{g'}}\sin^{-1} \;\; \sqrt{\frac{g'\tilde{r}}{v_1^2 + 2g'R}}\,\theta_f \\ & + \;\; \frac{2}{g'\sin\theta_f} \;\; \sqrt{v_1^2 + 2g'[r(1-\cos\theta_f) + h]} - \sqrt{v_1^2 + 2g'r(1-\cos\theta_f)} \;\; . \end{aligned} \tag{2.39}$$

Although the above expressions are lengthy, it is important to notice that the time needed to traverse either track is simply a function of the initial velocity of the ball and the measurable parameters of the track.

III. EXPERIMENTAL SETUP

The high road/low road track was built as part of a senior research project in addition to the accompanying theoretical treatment presented in the aforementioned section and the data analysis presented in the next. A pair of rails were fashioned so that a steel ball nested on them as illustrated in Fig 7. The rails were constructed from quarter-inch steel rod. The shape of the lower track was engineered by bending the two rails to match a template laid out on a board. Trusses were built and welded vertically to the rails in order to support the tracks. Spacers were then welded at roughly 7 in. intervals in order to maintain an equal spacing between the rail pairs. A ramp was welded to one end of each track, which allows for a variety of speeds to easily be obtained. This is accomplished by simply varying the initial height of the ball on the ramp. The base of the track consists of a straight 2 in.$\times$10 in.$\times$10 ft. long piece of lumber. For support, holes were drilled into the wood to receive the ends of the trusses. An image of the track is displayed in Fig. 1. Table I shows the quantitative specifications of the track.

Measurements of the initial and final speed of the ball and the time interval needed for the ball to traverse a given track were made via Pasco Photogates and a Data Studio interface system. A photogate works by measuring the time interval that an infrared light in a given gate is blocked. From this measurement, Data Studio can then easily calculate the average speed of the object as it moves through a gate by comparing this time interval to

Track parameters	Symbol	Value
Total Length	L	3.00 m
Upper Horizontal	l_1	0.28 m
Lower Horizontal	l_2	0.91 m
Radius of Curvature	R	0.10 m
Rise of Diagonal	h	0.09 m
Run of Diagonal	d	0.24 m
Deviation Angle	θ_f	0.37 rad
Radius of Ball	R_b	0.0135 m
Effective Radius of Ball	R_e	0.0093 m
Radius of Rail	R_r	0.0033 m
Distance Between Rails	w	0.0179 m
Effective Gravitational Acceleration	g'	5.29 m/s^2

TABLE I: Parameters of Track

the flag length, which here equates to the diameter of the ball. Photogates can also measure the time interval between the activation of two successive gates. A set of photogates was placed on each track for a total of four gates. One photogate was positioned at the start of each track and was used to obtain the initial speed of the ball. A second photogate was positioned at the end of each track. This second photogate yielded the final speed of the ball and allowed for a measurement of the total time interval needed for the ball to traverse a given track. The gates were carefully positioned to ensure that both sets spanned equal horizontal distances. It was also important that the gates were placed at the same height above the track so to be tripped by the same relative location of the ball. For an accurate flag length, the infrared light in the photogate obviously needed to align with the ball at its widest point.

The total drag due to the rotational friction of the ball-track interaction and the velocity-dependent air resistance results in a slowing of the ball as it traverses the track. Hence, the speed of the ball as measured by the second photogate can be significantly smaller than that measured by the first. For smaller speeds, this slowing plays a more significant fractional role. As our theoretical determination of the time difference, as summarized in Eqs. (2.37)-(2.39), neglected the dissipative effects of friction, our data is expected to deviate from the theoretical curve. To account for the dissipative effects, we use the average speed, as obtained from the measured initial and final speeds, as numerical input for v_1 in the theoretical curve displayed in Fig. 6. It should be noted that in the limit of vanishing frictional effects, the average speed and v_1 become equivalent.

IV. EXPERIMENTAL PROCEDURE & DISCUSSION

Data from the experiment was first collected on the low road. The desired range of speeds was probed by releasing the ball from rest at a chosen initial ramp height. The ball was then allowed to traverse the track. Data Studio displayed values for the initial and final speeds and the time interval necessary for the ball to traverse the track. From this data, the average speed could be calculated for a given time interval. We repeated several

low road track runs with the intent of covering the widest range of speeds possible, while obtaining proper representation of the intermediate speeds. The minimum speed recorded corresponded to the ball that just finished the track with vanishing speed; the maximum corresponded to that where the ball just remained in contact with the track. For speeds above this maximum, the 'hum' of the ball on the track would lessen as the ball traversed the upper constant radius of curvature section thus coming unattached. These data points were discarded as our theoretical analysis demands that the ball stay in contact with the track.

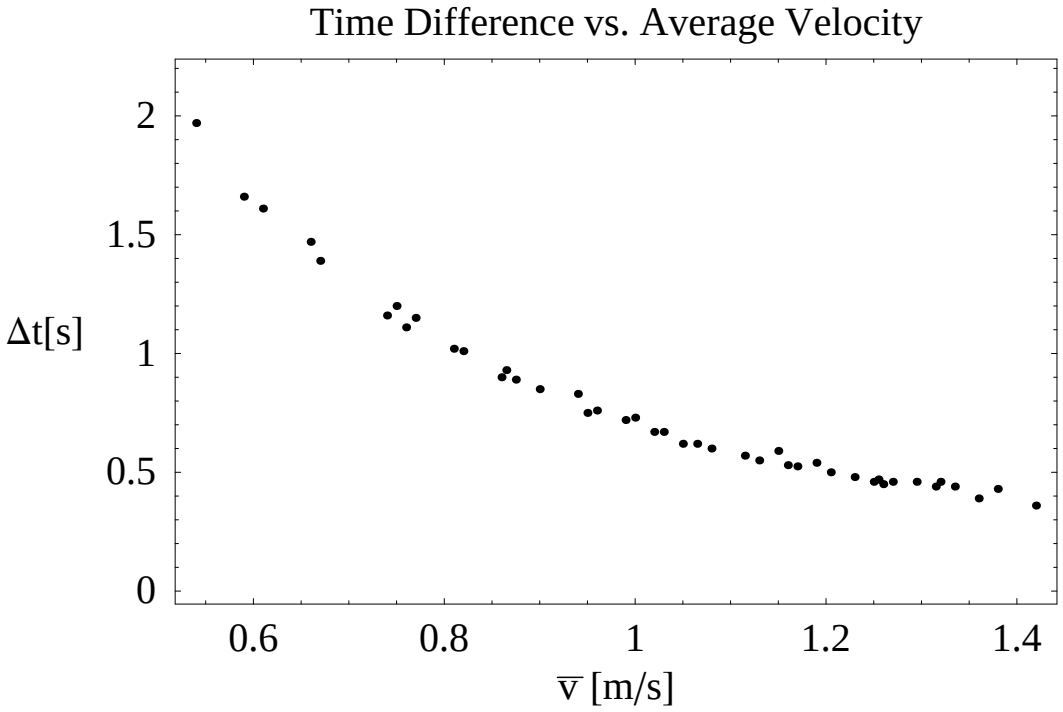


FIG. 6: Plot of the time difference, $\Delta t = t_H - t_L$, versus the average velocity of the ball, $\bar{v}$.

We then repeated the above process for the high road. The average speed for a given run was again calculated and compared to those of the low road. If this average speed matched any of those for runs on the low road (to within 0.005 m/s), then the time difference was calculated and the data point was recorded. Fig. 6 displays the data and the theoretical prediction, as calculated in section II, with the average speed used as numerical input for v_1.

In general, the high road/ low road demonstration serves as an excellent teaching tool of classical dynamics. The results of the demonstration are counterintuitive and can be used to engage the students who are often surprised of the outcome. A detailed theoretical analysis can proceed and be used to cover much of Newtonian mechanics. Finally, measurements of the time difference as a function of the average speed for a high road/ low road demonstration can easily be made in the laboratory and be used to help reinforce the concepts and theory

previously encountered.

[1] Leonard, Wiliam J. and Gerace, William J., The Power of Simple Reasoning, **Phys. Teach. 34**, 280-283, (1996)
[2] Thaden-Koch, Thomas Christian, A Coordination Class Analysis of College Students' Judgments about Animated Motion, *Ph.D. Dissertation*
[3] Rodgers, Eric, Brachistochrone and Tautochrone Curves for Rolling Bodies, **Am. J. Phys. 14** , 249-252 (1946)
[4] Venezian, Giulio, Terrestrial Brachistochrone, **Am. J. Phys. 34** , 701-704 (1966)
[5] Atanackovic, T.M., The Brachistochrone for a Material Point with Arbitrary Initial Velocity, **Am. J. Phys. 46** , 1274-1275 (1978)
[6] Stork, David G., Problem: Down in the Valley, **Am. J. Phys. 51** , 132, 174 (1983)
[7] Denman, Harry H., Remarks on Brachistochrone-Tautochrone Problems, **Am. J. Phys. 53** , 224-227 (1985)
[8] Stork, David G. and Yang, Ju-xing, The Unrestrained Brachistochrone, **Am. J. Phys. 54** , 992-997 (1986).
[9] Yang, Ju-xing, Stork, David G., and Galloway, David, The Rolling Unrestrained Brachistochrone, **Am. J. Phys. 55** , 844-847 (1987).
[10] Stork, David G., Problem: The Unrestrained Tachistos, **Am. J. Phys. 55** , No. 4, 296, 376 (1987)
[11] Fowler, Michael, Sliding Down a Cycloid, **Phys. Teach. 34**, 326, (1996)
[12] Edge, R.D., The Brachistochrone - or, the Longer Way Round May Be the Quickest Way Home, **Phys. Teach. 23**, 372-373 (1985)
[13] Tillotson, Kevin J., Who Will Win the Race, II?, **Phys. Teach. 28**, 537-538 ((1990)
[14] Figueroa, D., Gutierrez, G., and Fehr, C., Demonstrating the Brachistochrone and Tautochrone, **Phys. Teach. 35**, 494-498, (1997)
[15] Zheng, T.F., Gonci, K.A., Imparato, M.A., Ursulet, C., Donovan, K, Does Bead B Always Reach d First?, **Phys. Teach. 33**, 376-377 (1995)

APPENDIX A: EFFECTIVE RADIUS

To find the effective radius, R_e, we need to first calculate the angle of contact, α, between the ball and the rails. For an illustration of these quantities and the cross section of the ball and rails, see Fig. 7. As can be witnessed from the figure

$$2R_b \cos\alpha = w + 2R_r(1 - \cos\alpha) \tag{A1}$$

where w is the distance between the rails and R_b and R_r are the radii of the ball and rail, respectively. Solving for the angle and using basic trigonometry, one finds

$$\alpha = \cos^{-1}\left(\frac{w + 2R_r}{2(R_b + R_r)}\right) = \sin^{-1}\left(\sqrt{1 - \frac{1}{4}\left(\frac{w + 2R_r}{R_b + R_r}\right)^2}\right) \tag{A2}$$

In this study it is advantageous to find the ratio of the radius of the ball to the effective

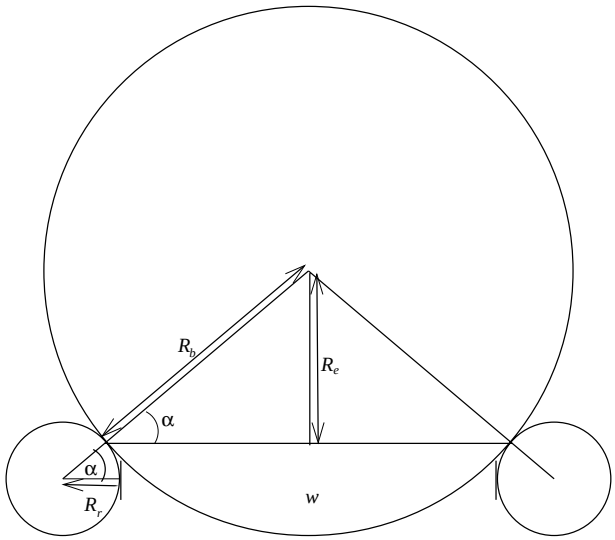


FIG. 7: Schematic of the cross section of the ball and the rails of the track.

radius, thus determining a unit-less scaling factor. Again from basic trigonometry, we find

$$\frac{R_b}{R_e} = \frac{1}{\sin\alpha} = \left(1 - \frac{1}{4}\left(\frac{w + 2R_r}{R_b + R_r}\right)^2\right)^{-1/2} = 1.46 \tag{A3}$$

to three significant figures where we used the data from Table I.

General Search for Stars with Rapid Optical Variations: Test Fields

E. Fagg, J. Park, K. Pearson and R. Kehoe
Southern Methodist University
Dallas, TX 75275

Key words: Astronomy, charged coupled devices (CCDs), eclipsing binary stars, light curves, magnitudes, pulsating stars, ROTSE telescopes, SIMBAD astronomical database, variable stars

Abstract:

We present a search for stars exhibiting short time-scale optical light variations. Our search employs archival data taken by the ROTSE1 telephoto array in two eight degree fields of view. This is a test study considering two fields which overlap fields previously mined for variables, but with different data and search techniques. Each field was observed for approximately six continuous hours. We employ a general search strategy based on statistical properties of the observed light curves for each object. The analysis is sensitive to sources with variations < 0.25 day and > 0.1 mag and with mean magnitudes between 9.5 mag and 14 mag. We identify 42 variable stars with our search strategy. Of these, 17 were not found by comparison with catalogs of previously acknowledged variables. Within this sample, attempts at classification yield four W UMa systems and two δ Scu stars. The remaining eleven transient detections exhibit incomplete light curves and require further study for classification.

I. Introduction

Stars with varying intensities of brightness, or "magnitude", also known as variable stars, have interested astronomers for hundreds of years. Intensity is "a measure of the light energy from a star that hits 1 square meter in 1 second[1].", and the magnitude is the brightness scale that is based on a constant intensity ratio, which was defined by Hipparchus. The brightest star, Sirius, is magnitude -1.43 and the dimmest star visible to the unaided eye is approximately magnitude 6. The intensity vs. time graph is termed a 'light curve', and variable stars have varying light curves.

Variation in stellar brightness was observed by several cultures, including the Chinese, who saw `guest stars' which were sometimes actual stars that were previously too dim to see by eye. In Western Europe, the first acknowledgement was with the sudden awareness to "new stars" (Latin 'novae stellae'), such as Tycho's star of 1572. Since then, thousands of variable stars have been discovered, and classified based on their peculiarities in variation. The observation of these celestial phenomena have allowed astronomers, both amateur and professional, to learn more about the night sky, particularly objects too dim to see with the unaided eye.

Variable stars are grouped into two major categories: intrinsic, those whose luminosity varies due to pulsating or other alterations in physical characteristics; and extrinsic, those who appear to vary in magnitude due to an eclipsing companion. An eclipsing binary or a variable star can be usually distinguished by distinctive, usual

patterns of the intensity versus time light curve. Long period variable stars have been studied for many years because their apparent brightness changes can be easily observed using a telescope. Their periods can last from weeks up to several years.

Short period variation has also been studied, but such stars can require more frequent observations to be well-measured. Such objects have periods that can last from less than an hour to a few days. These types of variables are more commonly the subject of professional research today because of their high energy outputs and patterns of light variation. Two examples of these types of variables are the RR Lyrae (period: 0.2 day ≤ T ≤ 1.0 day, normally 0.5 day) and δ Scuti stars (period: T ≤ 0.3 day, amplitude: 0.01 magnitude ≤ Δm ≤ 0.5 magnitude), which are pulsators and have typically lower luminosity variations and shorter periods than other variables. An example light curve for a pulsating variable is shown in Figure 1, which is discussed in more detail in Section IV. Such stars do not exhibit a static equilibrium between the outward pressure from radiation and the inward pull of gravity. Instead, they release energy when the zone expands and absorb energy when it is compressed, exhibiting complex alternations of heating and cooling, expansion and contraction, and growing more luminous and less luminous.

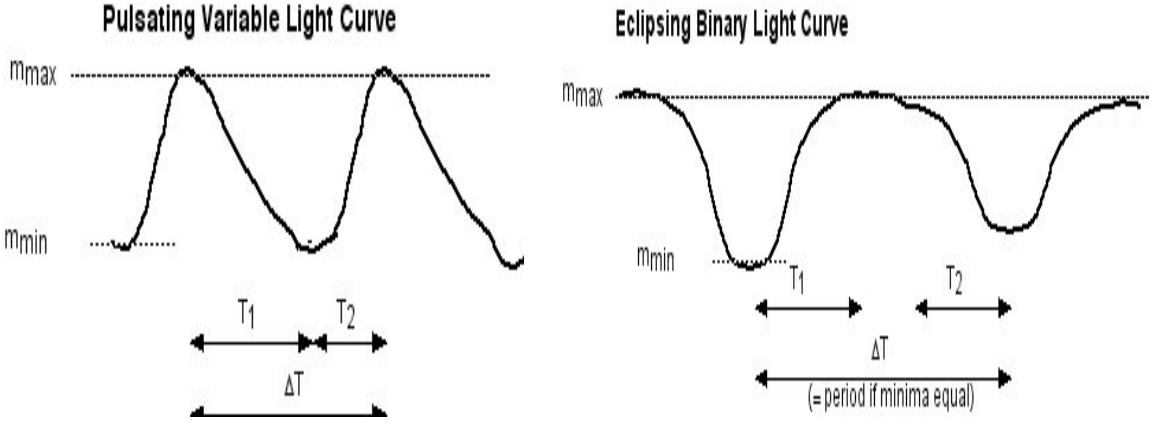


Figure 1: A pulsating variable light curve is shown on the left. The rise (T_1) and fall times (T_2) are generally not equal, and the total period (ΔT) equals the sum of these times. On the right is an eclipsing binary light curve. Dimming and brightening times are similar, and the total period (or half period) is longer than the sum of these times.

Eclipsing binary systems containing stars relatively close or touching each other also exhibit short period behavior. Some have light curves generally characterized by a span of brighter magnitude with one or two dips to a dimmer magnitude. A sketch of such a light curve is shown in Figure 1. Binary systems of the W UMa type contain stars that are so close that the surfaces are in contact and the stars share their atmospheres[2]. The second law of thermodynamics states that the entropy of an isolated system which is not in equilibrium will increase until equilibrium is attained. In a binary star system, heat transfers from the body of higher temperature to the body of lower temperature, therefore the heat, and thus luminosity, is transferred from the more massive star to the less massive one until equal temperatures are acquired. This model of stars is best represented by a light curve that resembles that of a typical binary system, with the dips in magnitude becoming less different from each other.

There are many other types of variable source that exhibit rapid optical light variation. Flare stars exhibit substantial changes in brightness in a few minutes before relaxing to quiescence. Cataclysmic variables can have erratic light curves. Other objects, not stars but active galaxies, can also exhibit rapidly varying light curves.

Many telescopes including the Robotic Optical Transient Search Experiment[3] (ROTSE) observe sources with short time-scale variation. The ROTSE1 telescope provided the data used in this paper. We have pursued a study in an unbiased way to identify the catalog of variable stars with gross characteristics of variation which do not assume a particular kind of physical system. We have employed a basic statistical analysis looking for light curves showing variation during much of a given night of observation.

II. Detector and Data

The ROTSE telescopes were originally created to study the optical light emitted by gamma ray bursts (GRBs), but they are also used to study optical light from numerous types of sources, including variable stars. These small but powerful telescopes are distributed and used internationally. We utilize a subset of data taken between 1999 and 2000 with the ROTSE1 telescope for the purpose of looking for isolated optical bursts associated with GRBs[4].

The ROTSE1 telescope was an array of four Canon telephoto lenses with 8 degree field of view each. These lenses were attached to Apogee AP-10 charge-coupled device (CCD) cameras. A CCD is an array of cells which send an electric signal based on the intensity of the light striking them. The AP-10 was designed initially for ROTSE1 and had a relatively high number of pixels (4 megapixels) and low electronics noise.

The ROTSE telescopes take measurements when photons emanating from a light source are focused to an image on the CCD array of cells. Typically in astronomy, filters are used to restrict to a standard range the wavelengths of light which can strike the CCD. A 'V band' filter restricts the signal to wavelengths in the middle of the optical spectrum, while an 'R band' filter restricts to redder wavelengths. In order to allow the maximum light to strike the CCD, the ROTSE optics do not include filters. The resulting ROTSE1-measured magnitude corresponded approximately to a V or R band magnitude.

The optics typically focus the signals from a point source such as a star to a region approximately 14 arc seconds wide. This is the size of an AP-10 CCD cell. For the brightest stars, signals can spill over to many neighboring cells. This is termed 'saturation' and it creates an undesirably bright and distorted image. As a result, it is difficult to obtain accurate photometry for stars brighter than 10^{th} mag in one minute exposures. Also, the low level electronics noise of the CCD makes it difficult to find or accurately measure the brightness of dim stars. Therefore, ROTSE1 is not sensitive to stars below approximately 15^{th} magnitude, since dim stars are more sensitive to the noise.

Many exposures of sixty seconds were obtained for several consecutive hours over several days, using different fields in April, 2000 and July, 2000 (and other periods), as part of a study for lone optical bursts[4]. We use the data for camera 'b' from the April 14 data, and camera 'a' for the July 6 data as test cases for a future analysis of the entire untriggered sample. These fields are termed '000414b' and '000706a', respectively. Approximately 6 consecutive hours of data were taken in each night. The July data was taken by pointing at two fields alternately in blocks of several minutes. This paper only

includes camera 'a' data for one of these two pointings. The images were processed through a chain which corrects for lens vignetting and noise effects, finds stars from the cell readouts in the CCD[5], and provides a precision calibration of positions and magnitudes by comparing to the Hipparcos astrometry catalog[6]. Typically, a position resolution of 1.5 arc seconds is achieved. For bright sources, a magnitude uncertainty of 0.02 magnitudes is achieved.

For each object, many observations can be obtained in one night. With the use of programs written at the University of Michigan using the Interactive Data Language (IDL), these collected data of magnitude are plotted versus time, creating a light curve.

III. Lightcurve Selection

Typically, the fields used for this analysis yield approximately 20,000 observed sources per image. The vast majority of these are non-variable objects, so a search must be made for varying lightcurves. The data stored in the University of Michigan's archive can be extracted through a special lightcurve search called "find_burst." Data is selected by quantifying search "cuts" input by the user based on three statistical criteria calculated in the ways described below. Two of the parameters account for the level of error present in each measurement in the lightcurve.

Amplitude of magnitude variation (Δm): We calculate the difference between the brightest observed magnitude and the dimmest for the light curve:

$$\Delta m = m_{max} - m_{min} . \quad (1)$$

We require Δm to be greater than some value.

Significance of maximum variation (σ_{max}): We calculate this as the difference in magnitude divided by the estimated errors on the magnitude measurement, added in quadrature:

$$\sigma_{max} = (m_{max} - m_{min}) / (\varepsilon_{max}^2 + \varepsilon_{min}^2)^{1/2} . \quad (2)$$

Here, m means the magnitudes of the light curve, and ε means estimated errors. These errors are statistical errors based on the brightness of the star. This parameter attempts to determine if an observed variation is significant by seeing if it is large compared to the errors on the measurements. We require σ_{max} to be greater than some value to select variable stars which have large, significant variations.

Chi-squared (χ^2): This is calculated as

$$\chi^2 = \sum ((m_i - m_{avg})/ \sigma_i)^2 \quad (3)$$

where the sum is performed over all measurements, i, and m_i and σ_i are the magnitude and estimated error for each measurement, respectively. We calculate m_{avg} to be the average magnitude for the whole light curve. In calculating χ^2, we exclude the single measurement furthest from m_{avg}.

This rejects fake variable light curves where just one measurement is bad. If the star was not variable, its χ^2 would be near 1.0. We require the light curve to be inconsistent with constant brightness, so χ^2 should be large.

To perform a general search, we tried several combinations of cuts on the 000414b data using the above three parameters and counted the number of objects which fell into four general categories. The brightest sources exhibit light curves where most or all of their measurements are saturated, which results in a highly discontinuous light curve, which is not a sign of a variable star. We omit the candidates from our list that are brighter than the magnitude in which this phenomenon is no longer observed. The dimmest sources are strongly affected by noise in the camera and often may not even be identified, resulting in a sparse and erratic light curve with poorly assessed errors on the measurements. We omit stars dimmer than the brightest stars in which we begin observing this phenomenon.

Two types of light curves are found among the remaining candidates. The first type exhibits at least two abrupt, consecutive and opposite changes in brightness, producing one observation substantially different than immediately preceding and following measurements. Usually there must be two such occurrences because the χ^2 requirement rejects candidates with just one isolated bad observation. The other type exhibits a light curve in which most observations are continuous with their immediate neighbors in time. This is not to say they have regular variation: such light curves can be very variable on the timescale of several minutes.

Although it is very difficult, based on general physics principles, to have objects fluctuate by > 0.1 magnitude in brightness in one minute, extreme examples such as gamma-ray bursts have been observed. We keep our search general, while also yielding a purer sample of clear variables, by taking the following approach. We visually inspect the light curves and omit all observations that are discontinuous from their preceding and succeeding observations. In other words, we permit an object to change by several times its error in 1 minute, or we allow it to dim similarly, but we do not permit it to brighten and return back to its original state in three consecutive observations. Such variations can happen rarely due to improperly corrected or noisy pixels, but can produce a large number of false candidates because we are analyzing over 330 observations for 20,000 stars in each field. In making this requirement, we remove isolated bad observations while remaining sensitive to variation on the timescale of a few minutes without bias.

The results of our scan of selections, with the number of rejected bright or dim candidates, and those from bad observations, are shown in Table 1.

Δm	σ_{max}	χ^2	bright	bad obs.	Dim	candidates
0.5	2.0	3.0	0	84	117	2
0.1	0.0	3.0 (‘B’)	24	63	171	15
0.1	0.0	5.0	18	7	15	11
0.1	5.0	1.0 (‘A’)	54	410	195	16
1.0	2.0	5.0	18	0	0	0
1.0	5.0	0.0	0	30	27	0

Table 1: Cuts used in initial lightcurve search in a test on field 000414b. The four columns on the right give the number of backgrounds of different types, and the

number of good candidates. Cuts that are labeled 'A' and 'B', which have no relation to camera names 'a' and 'b', were chosen for variable star selections in this paper.

The rightmost column indicates the number of good variable candidates we observe. The results support the following conclusions. Cuts using $\Delta m > 0.1$ yielded the most good variable candidates: i.e. $\Delta m = 0.1$, $\sigma_{max} = 5.0$, $\chi^2 = 1.0$ yielded sixteen variables out of over 670 light curves. The use of the χ^2 cut significantly reduces the background due to bad observations, with a value of 3.0 (selection 'B') showing the next largest number of good candidate light curves. A selection emphasizing the χ^2 parameter yields a lower background with a similar number of good objects. For field 000414b, selections 'A' and 'B' find a common set of eleven variables candidates, plus five with 'A' only and four with 'B' only.

IV. List of Objects

A total of forty-two candidates are identified in the two fields. For the 000706a field, the 22 identified candidates were varying in at least one other night. All candidates are listed in Table 2 and Table 3, along with their positions and measured lightcurve characteristics. The designations at left follow the form 'ROTSE1 JHHMMSS.SS±ddmmss.s' where 'HH', 'MM' and 'SS.SS' give the RA position in hours, minutes and seconds, and the 'dd', 'mm' and 'ss.s' give the declination analogously. The ROTSE1 magnitude, m_{R1}, is calculated as the average of m_{max} and m_{min}. The time interval ΔT is calculated as the apparent time from maximum to maximum, or minimum to minimum, if a pair of either is apparent. In the case of δ Scu and other pulsating stars, this will be the period. For symmetric W UMa stars and other eclipsing binaries where both dips are deep, it will be half the period. For longer period variables, the quoted value is merely a lower limit as only 6 hours of data were used. Because of the length of time ROTSE1 was observing in the night, there is a natural insensitivity to medium or long time-scale variables.

Also shown in Table 2 and Table 3 are the rise time and fall time of the lightcurve, labeled T_1 and T_2, respectively. These are shown graphically in Figure 1. T_1 is defined as the *shortest* time it takes to rise from the end of a period of minimum brightness to the beginning of a period of maximum brightness. T_2 is defined conversely. These parameters may not correspond to the *full* time to transition from minimum to maximum brightness for the star. But if m_{min} and m_{max} correspond to the actual minima and maxima, these parameters assist in classification efforts, which will be described later.

Source	m_{R1}	Δm	ΔT (days)	T_1(days)	T_2(days)
ROTSE1 J110448.11+353626.6	10.96	0.21	0.20	0.09	0.09
ROTSE1 J113721.25+425544.6	11.49	0.36	0.21	0.10	0.09
ROTSE1 J111345.07+423951.7	11.77	0.15	>0.25	0.16	0.15
ROTSE1 J113334.68+425829.2	11.81	0.28	>0.25	0.11	0.19
ROTSE1 J113536.72+384557.5	12.01	0.47	0.25	0.09	0.20
ROTSE1 J112141.02+433653.1	11.70	0.28	0.25	0.12	0.13
ROTSE1 J111719.74+394303.0	11.98	0.31	0.10	0.08	0.07
ROTSE1 J112541.63+423448.8	12.11	0.35	0.18	0.08	0.09
ROTSE1 J111305.98+402137.7	11.89	0.36	>0.25	0.12	0.17

ROTSE1 J112037.62+392100.3	12.15	0.12	0.07	0.03	0.04
ROTSE1 J111615.06+355027.2	12.66	0.39	0.20	0.09	0.11
ROTSE1 J113928.29+403632.8	12.85	0.50	0.25	0.13	0.12
ROTSE1 J111340.03+424413.8	12.68	0.26	0.21	0.09	0.13
ROTSE1 J111415.57+371825.6	13.02	0.10	0.05	0.02	0.02
ROTSE1 J111105.45+381123.5	13.22	0.21	0.08	0.03	0.03
ROTSE1 J111716.02+385716.9	13.24	0.27	0.20	0.08	0.10
ROTSE1 J112148.80+405938.4	13.25	0.26	0.21	0.07	0.04
ROTSE1 J112009.02+435349.0	11.52	0.17	0.25	0.14	0.12
ROTSE1 J111734.08+410649.0	13.45	0.50	0.10	0.06	0.05
ROTSE1 J110340.78+402617.1	13.80	0.60	0.11	0.07	0.03

Table 2: Candidate variable stars identified by light curve selection in field 000414b. Light-curve properties are tabulated.

Source	m_{R1}	Δm	ΔT (days)	T_1(days)	T_2(days)
ROTSE1 J154029.81+453200.6	13.25	0.39	0.15	0.07	0.05
ROTSE1 J154136.92+515926.5	13.42	0.42	0.075	0.02	0.055
ROTSE1 J154436.45+461922.0	12.54	0.32	0.12	0.04	0.08
ROTSE1 J155028.43+455751.0	12.83	0.29	0.065	0.015	0.05
ROTSE1 J155600.54+494757.0	13.05	0.36	>0.24	0.07	0.11
ROTSE1 J155705.19+500527.6	12.48	0.45	0.16	0.08	0.06
ROTSE1 J155809.25+485742.7	12.58	0.55	>0.24	0.12	--
ROTSE1 J155825.31+492652.1	11.88	0.15	0.16	0.08	0.045
ROTSE1 J155853.75+463548.7	13.34	0.44	0.12	--	0.075
ROTSE1 J160032.12+465526.8	11.70	0.36	>0.17	0.06	0.11
ROTSE1 J160048.24+511648.0	13.32	0.30	0.13	0.06	0.06
ROTSE1 J160121.91+482938.3	12.70	0.55	0.15	0.06	0.06
ROTSE1 J160434.17+504514.5	12.56	0.23	0.18	0.09	0.08
ROTSE1 J160602.27+501111.4	10.23	0.51	0.21	0.08	0.08
ROTSE1 J160653.63+513835.7	13.44	0.67	0.13	0.06	0.07
ROTSE1 J161033.65+514401.1	11.20	0.16	0.20	0.09	0.10
ROTSE1 J161134.29+471612.6	13.52	0.62	0.14	0.05	0.06
ROTSE1 J161321.76+515524.2	11.51	0.45	0.16	0.09	0.07
ROTSE1 J161506.00+445822.3	12.11	0.30	>0.17	0.06	0.08
ROTSE1 J161801.01+511153.0	13.83	0.65	>0.22	0.07	--
ROTSE1 J162004.25+451259.4	9.62	0.07	>0.23	0.23	--
ROTSE1 J162410.37+455527.0	10.29	0.42	0.13	0.06	0.06

Table 3: Candidate variable stars identified by lightcurve selection in field 000706a. Lightcurve properties are tabulated.

Source	Nearest Object	Δr (arcsec)	Object type
ROTSE1 J110448.11+353626.6	HH UMa	0.51	Contact binary[6]
ROTSE1 J113721.25+425544.6	[GGM 2006] 4974567	4.86	Contact binary[7]
ROTSE1 J111345.07+423951.7	TYC 3012-1895-1	0.43	Star
ROTSE1 J113334.68+425829.2	MT UMa	0.30	W UMa[8]
ROTSE1 J113536.72+384557.5	MU UMa	0.55	RR Lyr[8]
ROTSE1 J112141.02+433653.1	MQ UMa	0.87	W UMa[8]
ROTSE1 J111719.74+394303.0	FIRST J111722.9+394253	38.28	Radio[9]
ROTSE1 J112541.63+423448.8	BS UMa	0.80	Algol[10]
ROTSE1 J111305.98+402137.7	MO UMa	0.38	RR Lyr[8]

ROTSE1 J112037.62+392100.3	MP UMa	0.15	Pulsating[4]
ROTSE1 J111615.06+355027.2	[GGM 2006] 7575961	0.13	contact binary[7]
ROTSE1 J113928.29+403632.8	FIRST J113922.2+403640	68.63	Radio[11]
ROTSE1 J111340.03+424413.8	---	---	---
ROTSE1 J111415.57+371825.6	---	---	---
ROTSE1 J111105.45+381123.5	SDSS J111055.84+381055.1	116.74	Quasar[12]
ROTSE1 J111716.02+385716.9	---	---	---
ROTSE1 J112148.80+405938.4	FIRST J112148.9+405909	29.15	Radio[11]
ROTSE1 J112009.02+435349.0	GB6 B1117+4411	99.11	Radio[13]
ROTSE1 J111734.08+410649.0	FIRST J111740.0+410628	69.94	Radio[11]
ROTSE1 J110340.78+402617.1	---	---	---

Table 4: Nearest matches in SIMBAD database to observed candidate variables from the 000414b field. Notes on object classification are given in the rightmost column.

Source	Nearest Object	Δ*r* (arcsec)	Object type
ROTSE1 J154029.81+453200.6	---	---	---
ROTSE1 J154136.92+515926.5	---	---	---
ROTSE1 J154436.45+461922.0	TYC 3483-746-1	1.58	δ Scu[14]
ROTSE1 J155028.43+455751.0	TYC 3490-814-1	0.87	δ Scu[14]
ROTSE1 J155600.54+494757.0	BPS BS 16029-0008	3.92	Star[15]
ROTSE1 J155705.19+500527.6	[GGM2006] 5207106	1.29	contact binary[7]
ROTSE1 J155809.25+485742.7	---	---	---
ROTSE1 J155825.31+492652.1	V* V1023 Her	1.10	W UMa
ROTSE1 J155853.75+463548.7	---	---	---
ROTSE1 J160032.12+465526.8	V* AR Her	1.57	RR Lyr
ROTSE1 J160048.24+511648.0	GSC 03497-01775	2.67	W UMa[16]
ROTSE1 J160121.91+482938.3	[GGM2006] 5208621	0.78	contact binary[7]
ROTSE1 J160434.17+504514.5	GSC 03497-00900	6.64	W UMa[16]
ROTSE1 J160602.27+501111.4	V* V842 Her	1.78	W UMa
ROTSE1 J160653.63+513835.7	---	---	---
ROTSE1 J161033.65+514401.1	TYC 3497-1342-1	0.44	Star[17]
ROTSE1 J161134.29+471612.6	GSC 03491-00010	0.91	W UMa[16]
ROTSE1 J161321.76+515524.2	[GGM2006] 5214166	2.94	contact binary[7]
ROTSE1 J161506.00+445822.3	---	---	---
ROTSE1 J161801.01+511153.0	GSC 03498-01093	1.10	RR Lyr[16]
ROTSE1 J162004.25+451259.4	V* V893 Her	1.91	RR Lyr
ROTSE1 J162410.37+455527.0	TYC 3492-1272-1	0.73	Star[18]

Table 5: Nearest matches in SIMBAD database to observed candidate variables for the 000706a field. Notes on variable classification are in the rightmost column.

Catalog Matching

We have compared these lists to existing objects catalogued in the SIMBAD[19] astronomical database according to their right ascension and declination. Close matches to known or suspected variable stars in the catalog were found for twenty-two candidates. We further check the AAVSO site[20] for remaining objects and find three more matches. All 25 matched candidates proved to be either pulsating variables or eclipsing binaries. Most of these were identified using ROTSE1 data in previous publications. Of the 25 matched candidates, twelve are recent, unconfirmed identifications. We present these

lightcurves in this section. From the previously acknowledged twenty-five, we have seventeen other candidates that are newly identified transients. We discuss these seventeen cases in the next section.

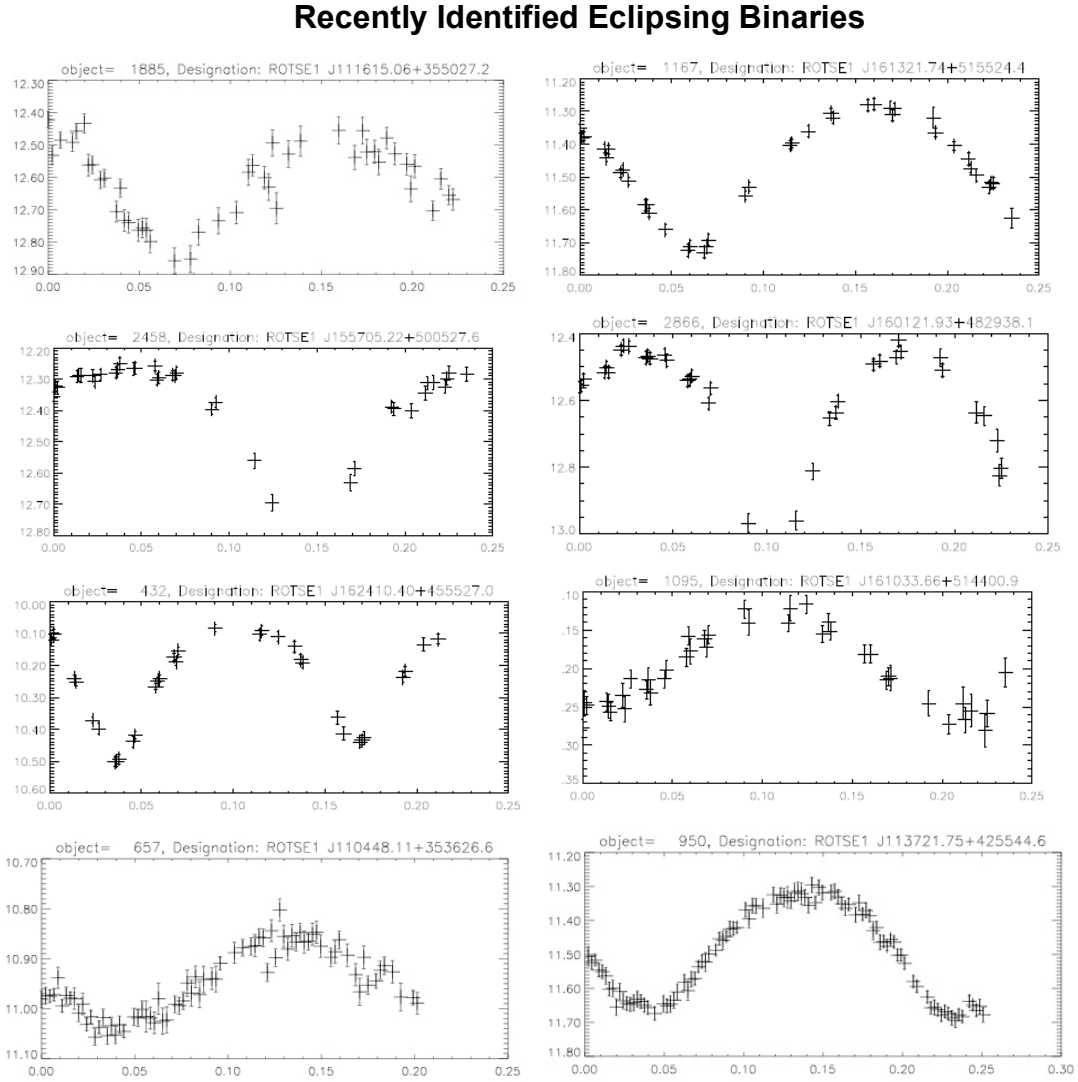


Figure 2: Eight single-night light curves for previously suspected contact binaries. Plotted errors are statistical + systematic. The *x*-axis represents the time of observation in days, while the *y*-axis represents the magnitude of an observation.

From our original 25 previously identified candidates, there are sixteen candidates matching previously identified or suspected eclipsing systems. Eight of the sixteen candidates are not listed as definite classifications in the SIMBAD catalog, although there are claims these are W UMa systems[6,7]. The light curves for the eight suggested variables are shown in Figure 2. We measure T_1 to be similar to T_2 and ΔT to be a little larger than T_1+T_2 for these objects, which suggest W UMa classification. These light curve data therefore support the existing claims.

Recently Identified Pulsating Variables

The remaining nine of twenty-five previously identified candidates are identified as pulsating variables. Four of them are not clearly categorized in SIMBAD. Their lightcurves are shown in Figure 3. One of the four candidates (ROTSE1 J112037.63+392100.3) is listed as a variable of general 'pulsating' classification in the SIMBAD catalog. This is based on a preliminary identification of this as a δ Scu star[4]. Three and a half more periods are evident in the lightcurve in Figure 3. At least two periods are evident for two of the other candidates. Considering that $T_1 \neq T_2$ and $\Delta T = T_1 + T_2$, these data support the claims of these being of δ Scu type. The last of the four stars is an RR Lyr star for which we observe a partial lightcurve consistent with this classification.

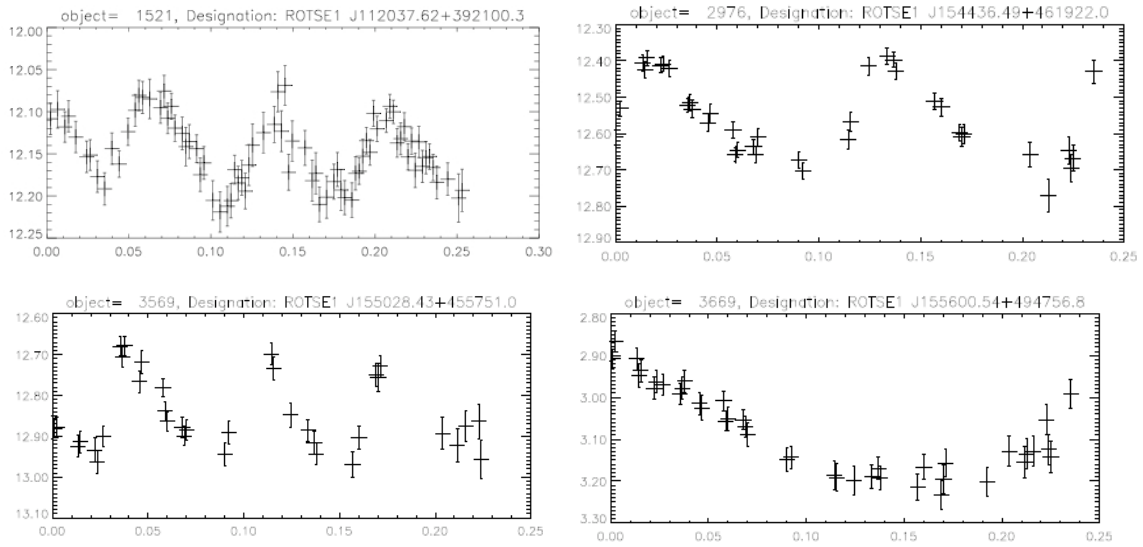


Figure 3: Four single-night light curves for previously suspected pulsating variables. Plotted errors are statistical + systematic. The *x*-axis represents the time of observation in days, while the *y*-axis represents the magnitude of an observation.

V. Analysis of Unmatched Sources

Our seventeen remaining candidates are previously uncatalogued as variable sources. We divide these variables into four categories: eclipsing candidates, pulsating candidates, incomplete smooth lightcurves, and irregular variables. This is based upon the lightcurve characteristics in Table 2 and Table 3. In order to assist identification of the candidates, it is useful to examine multiband photometry of these objects. In order to do this, we consulted the IPAC[21] online catalog to extract infrared or optical band magnitudes. For field 000414b, one of the candidates matching a radio source, ROTSE1 J112148.80+405938.4, has been suggested to be a variable of potentially pulsating type[4]. The lightcurve presented here does not support the pulsating classification, but it does confirm the variable nature of the source.

Previously Uncatalogued Eclipsing Candidates

Eclipsing candidates are identified by exhibiting clear minima and maxima which

allow a full set of light curve parameters to be calculated in Table 2 and Table 3. We select those variables where T_1 and T_2 are similar in magnitude, and where $\Delta T > T_1+T_2$. Their light curves are shown in Figure 4.

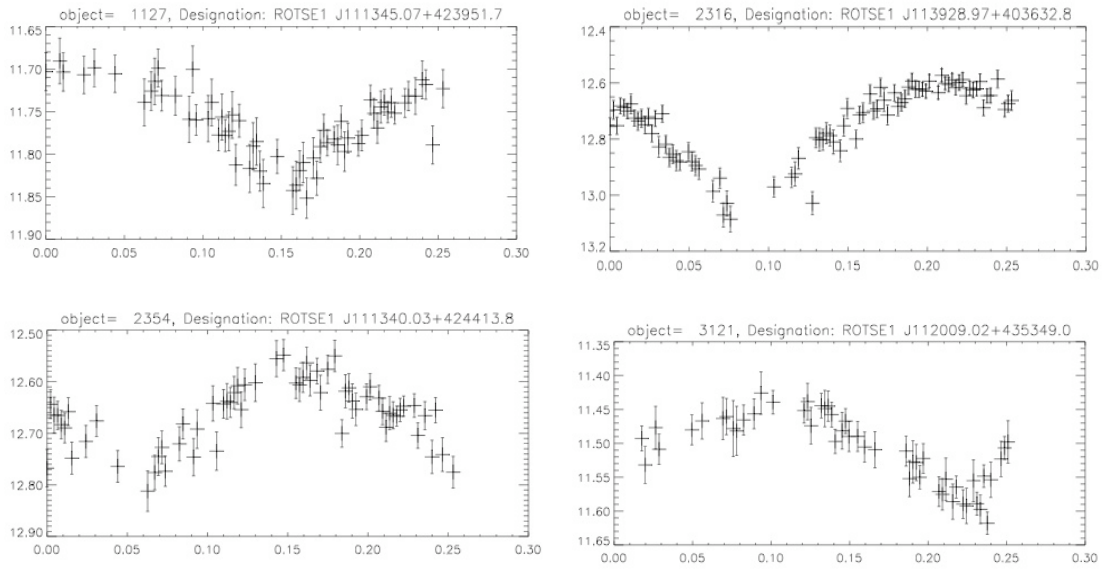


Figure 4: Single-night light curves for eclipsing binary candidates without a known variable match. Plotted errors are statistical + systematic. The *x*-axis represents the time of observation in days, while the *y*-axis represents the magnitude of an observation.

The estimated periods and infrared band magnitudes for these candidates are given in Table 6. The infrared band magnitudes J, H and K cover ranges of progressively longer wavelengths, with J being near-infrared and K being far-infrared wavelengths. All four objects satisfy the selections used in Ref. 7 to identify W UMa stars: $0.26 < T < 0.6$, m_{R1}-J ≤ 3.0, H-K ≤ 0.35, and $0.71\text{-}1.45T <$ J-H $< 0.96\text{-}1.45T$. The amplitudes of variation for these stars tend to be lower than was typical in that search.

Source	T (days)	J	H	K	m_{R1}-J	H-K	J-H
ROTSE1 J111345.07+423951.7	0.55	10.925	10.735	10.660	0.84	0.08	0.19
ROTSE1 J113928.29+403632.8	0.50	11.940	11.697	11.625	0.91	0.07	0.24
ROTSE1 J111340.03+424413.8	0.42	11.422	11.165	11.103	1.26	0.07	0.25
ROTSE1 J112009.02+435349.0	0.5	10.427	10.234	10.147	1.09	0.08	0.20

Table 6: Period, infrared band magnitudes and colors for four new W UMa candidates. J through K bands are from 2MASS[22].

Previously Uncatalogued Pulsating Candidates

The pulsating candidates are identified by exhibiting minima and maxima. These candidates satisfy $T_1 \neq T_2$ and ΔT approximately equal to T_1+T_2. Their lightcurves are shown in Figure 5. ROTSE1 J111415.57+371825.6 is not a clear variable detection. The second half of the lightcurve is very erratic, suggesting an instrumental problem, while the earlier 3 hours may indicate a δ Scu with ~0.1 mag amplitude and 0.05 day period at the edge of sensitivity. A firm identification will require examining more data.

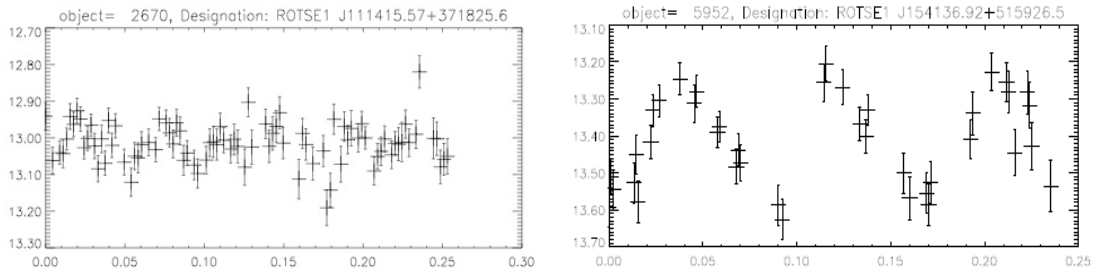


Figure 5: Single-night light curves for objects without a match but which may be pulsating variables. Plotted errors are statistical + systematic. The *x*-axis represents the time of observation in days, while the *y*-axis represents the magnitude of an observation.

Color information for ROTSE1 J154136.92+515926.5 and ROTSE1 J111415.57+371825.6 is given in Table 7. Pulsators of δ Scu type are expected to reside in spectral categories A0-F5, which corresponds roughly to $-0.1 < B\text{-}V < 1.0$. We do not have B and V magnitudes for these two objects, but obtain B and R by averaging the two measurements from the USNO-B1 catalog. If we assume that B>V and V>R, then B-R is an upper limit on B-V. These quantities are consistent with both stars being δ Scu stars.

Source	*T* (days)	B	R	B-R
ROTSE1 J111415.57+371825.6	0.05	13.68	13.00	0.68
ROTSE1 J154136.92+515926.5	0.08	13.39	12.67	0.72

Table 7: Period, optical band magnitudes and color for two δ Scu candidates. B and R band magnitudes are from USNO-B1.

Previously Uncatalogued Stars with Continuous, Incomplete Variations

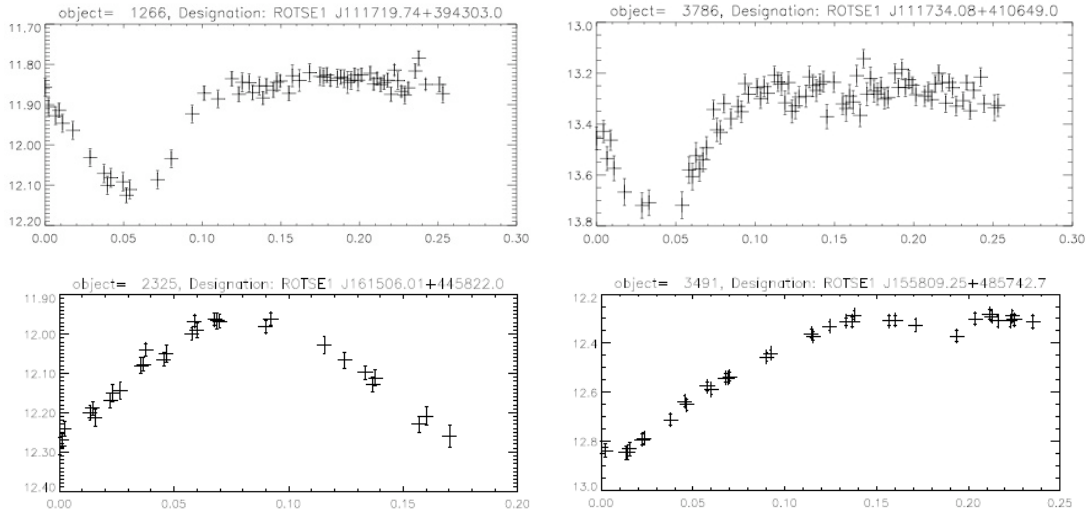


Figure 6: Single-night light curves for objects without a transient match. Lightcurves appear to be incomplete segments from regular variables. Errors are statistical + systematic. The *x*-axis represents the time of observation in days, while the

y-axis represents the magnitude of an observation.

Four variables with incomplete smooth lightcurves are shown in Figure 6. These lightcurves are missing some information needed for a preliminary classification. In all cases, we cannot properly estimate T_2 or ΔT, and either the maxima or minima are missing. Although they appear clearly distinct from the irregular variable category, use of further data will be needed to establish these identifications.

Previously Uncatalogued Irregular Variable Candidates:

Seven of the candidates appear to fall into an irregular variable category. These lightcurves are shown in Figure 7. We will present further results on these objects in a future study.

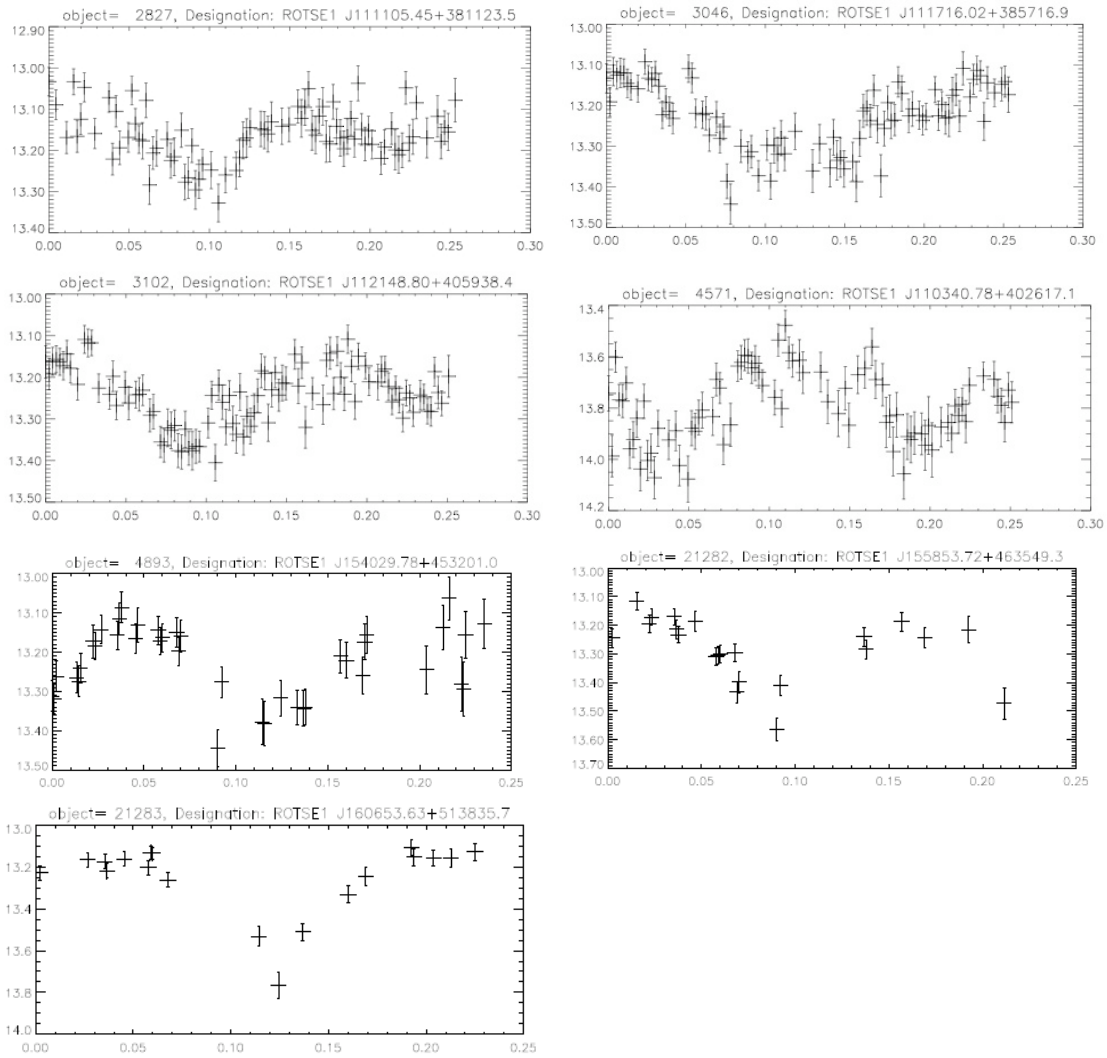


Figure 7: Single-night light curves for objects without a variable match and exhibiting irregular variations. Plotted errors are statistical + systematic. The *x*-axis represents the time of observation in days, while the *y*-axis represents the magnitude of an observation.

VI. Results

We have performed a search for optically variable sources in two fields with one night each of ROTSE1 data and identified 42 variable candidates. Of these, 10% were previously discovered with non-ROTSE1 data, and 17 were not previously catalogued as variables. We provide light curves to support classification for twelve W UMa, δ Scu and RR Lyr variables whose identification may not be settled. We identify four previously unidentified W UMa contact binaries and two δ Scu stars. We also identify 11 candidates that are clearly variable but require more data to ascertain their physical characteristics.

This work was supported by the Lightner-Sams Foundation grant to the Department of Physics at SMU, as well as matching funds from the SMU Undergraduate Research Associate Program. We thank the ROTSE Collaboration and the U. of Michigan group for their assistance, with particular thanks to Fang Yuan and Heather Swan for technical assistance, and Professors Carl Akerlof and Tim McKay for suggestions and access to the Michigan data and software archive for ROTSE1. This research made use of the SIMBAD database operated at CDS, Strasbourg, France. We also appreciate the use of the IPAC online service to provide access to 2MASS and USNO measurements.

References:

[1] M. A. Seeds, Foundations of Astronomy.
[2] C. Sterken and C. Jaschek, Light Curves of Variable Stars: A Pictorial Atlas. Cambridge University Press, 1996.
[3] http://www.rotse.net/
[4] R. Kehoe, et al., *Astrophys. J.*, 577:845 (2002).
[5] E. Bertin and S. Arnouts, *Astron. Astrophys.,* 117:393 (1996).
[6] M. Perryman, et al., Hipparcos Catalogue, *Astron. Astrophys.*, 323:L49 (1997).
[7] S. Gettel, M. Geske, and T. McKay, *Astron. J.*, 131:621 (2006).
[8] E. Kazarovetz, E. Pastukhova, and N. Samus, *Perem. Zvezdy*, 25:2 (2005).
[9] R. White, et al., *Astrophys. J.*, 475:479 (1997).
[10] A. Norton, et al., *Astron. Astrophys.*, 467:785 (2007).
[11] R. Becker, R. White, and D. Helfand, *Astrophys. J.*, 450:559 (1995).
[12] D. Schneider, et al., *Astron. J.,* 134:102 (2007).
[13] J. Douglas, et al., *Astron. J.*, 111:1945 (1996).
[14] A. Khruslov, *Perem. Zvezdy,* 6:7 (2006).
[15] P. Wils, C. Lloyd and K. Bernhard, *Mon. Not. R. Astron. Soc.*, 368:1757 (2006).
[16] A. Khruslov, *Perem. Zvezdy*, 6:16 (2006).
[17] A. Khruslov, *Perem. Zvezdy*, 7:6 (2007).
[18] A. Khruslov, *Perem. Zvezdy*, 6:20 (2006).
[19] http://simbad.harvard.edu/simbad/
[20] http://www.aavso.org/vsx
[21] http://irsa.ipac.caltech.edu/
[22] R. Cutri, et al., *Collection of Electr. Catalogues*, 2246:0 (2003).

Journal of Undergraduate Research in Physics
August 31, 2009

Kayaking Physics: The Tipping Angle

Daniel R. Rottinghaus and Chad A. Middleton

Department of Physical and Environmental Sciences, Mesa State College,
Grand Junction, CO 81501

Email undergraduate researcher: drotting@mesastate.edu
Email research advisor: chmiddle@mesastate.edu

Abstract

We apply basic physics to the rotational motion of a kayak in water to determine the angle at which the kayak will tip. Approximating the cross-section of the kayak as an ellipse, we apply Newtonian mechanics to the physical system and arrive at a set of coupled algebraic equations describing this critical angle. The equations are transcendental in nature and cannot be solved analytically. Using a numerical approach, we calculate the tipping angle for several values of the center of mass of the kayak-kayaker system.

Key words: buoyancy, center of mass, fluid mechanics, Newtonian mechanics, torques, transcendental equations

Introduction

The dynamics of kayaking are governed entirely by basic physical properties. Anyone who has partaken in the joy of kayaking knows that if they lean too much to one side or the other, they will capsize the kayak. Our goal here is to determine the angle at which a kayak can balance on edge. This angle

corresponds to an unstable equilibrium point and will be referred to as the critical angle throughout the rest of this manuscript.

Tipping a kayak on its side may seem trivial, but the physics that dictates the dynamics is rather sophisticated. In fact, it is first necessary to simplify the problem with several assumptions in order to arrive at a manageable physical model. Here, we will treat the kayak and kayaker as a rigid body. We further simplify the model by analyzing the two-dimensional cross-section of the kayak and approximate the cross-section of the kayak to be that of an ellipse. This assumption is justifiable in that, although kayaks come in many different shapes, most of these shapes are slight variations of an elliptical cross-section. Treating the kayak as having an elliptical cross-section additionally allows for an easy consistency check of our equations. By setting the semi-major and minor axes to be of equal length, we can easily compare our equations of motion to that of an object of circular cross-section.

We begin by deriving the equations of a kayak in mechanical equilibrium by setting the net force and net torque acting on the kayak equal to zero. This yields a set of algebraic equations describing the critical angle in terms of the parameters of the system and the height of the water level relative to the center of the kayak. We then numerically solve for the critical angle for several values of the center of mass of the kayak-kayaker system and plot these results. The

resulting plot reveals the dependence of the critical angle on the location of the center of mass of the system.

The Physics of Kayaks

It is instructive to first consider an upright kayak in water in the absence of rotation**.** Essentially, the hull of the kayak is sitting with the semi-major axis parallel to the surface of the water. Figure 1 depicts this scenario.

In equilibrium, when the boat undergoes zero acceleration, the net force acting on the kayak is zero. The forces acting on the kayak are the force due to gravity, F_g , and the force due to buoyancy, F_B . From basic physics, the buoyant force is equal to the weight of the water displaced, or equivalently, the product of the density of water, the volume of water displaced, and the constant gravitational field. We set the buoyant force equal to the force due to gravity yielding

$$mg = \rho V g \tag{1}$$

where m is the mass of the kayak/kayaker system and V is the volume of the fluid displaced. To further simplify the problem, we demand that the cross-sectional area of the kayak is uniform throughout the length of the boat. Using this assumption, the volume of the water displaced by the kayak can be related to the cross-sectional area of the water displaced and the length of the kayak, l, and is given by

$$V = Al \tag{2}$$

Inserting (2) into (1) and solving for the cross-sectional area of the water displaced, we obtain

$$A = \frac{m}{\rho l} \tag{3}$$

Notice that the area of water displaced is a constant for the kayak and is independent of the angle of orientation. This constant area is determined by the mass of the kayak-kayaker system, the density of water, and the kayak's length.

Next, we rotate the kayak about an angle, θ, in the clockwise direction with respect to the initial coordinate system. Figure 2 shows this rotation. By considering the rotation of the system about the kayak's center of mass, the location of the center of mass of the displaced water becomes a dynamical quantity. When the center of mass of the kayak-kayaker system and the center of buoyancy[3] share the same x-coordinate, the net torque acting on the kayak is zero and the kayak is in an unstable equilibrium position. This scenario illustrates a second equilibrium position (in addition to the upright stable equilibrium position illustrated in Figure 1) when the kayak is tipped at this specific critical angle. When the centers of mass do not share the same x-coordinate, there exists a net torque acting on the system as shown in Figure 3. This causes the kayak to either right itself or to tip over.

In order to find this critical angle, we consider a rotation of the kayak about the center of elliptical cross-section. This is accomplished by using a rotation matrix[1].

$$\begin{pmatrix} y' \\ x' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} \tag{4}$$

When the multiplication is carried out, the end result relates the two coordinate systems.

$$y' = y\cos\theta + x\sin\theta \tag{5}$$

$$x' = x\cos\theta - y\sin\theta \tag{6}$$

To find the equation describing the rotated ellipse, we place these transformations into the equation of the ellipse

$$\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1 \tag{7}$$

where a and b are the semi-major and minor axes, respectively. Solving for the dependent variable, we obtain the expression

$$y_{\pm}(x) = \alpha\left(-\beta\ x\ \pm\sqrt{\frac{ab}{\alpha} - x^2}\right) \tag{8}$$

where α and β are functions of θ and are defined as:

$$\alpha \equiv \frac{ab}{a^2 - \sin^2\theta\left(a^2 - b^2\right)} \tag{9}$$

$$\beta \equiv \frac{\sin\theta\cos\theta\left(a^2 - b^2\right)}{ab} \tag{10}$$

Notice that the two solutions expressed in (8) arise from the fact that the ellipse is a multi-valued function. $y_+(x)$, $y_-(x)$ describes the upper, lower section of the rotated ellipse, respectively.

Now that an equation describing the ellipse in terms of its rotated orientation has been found, we proceed in finding an equation describing the center of buoyancy. The center of mass of a system bounded by two functions in the *xy* plane is [2]

$$\overline{X}_w = \frac{1}{A}\int_{XL}^{XR} x \quad \left[f(x) - g(x)\right]dx \tag{11}$$

where *A* is the area of the displaced water and X_L and X_R are the *x*-coordinates corresponding to where the two functions *f(x)* and *g(x)* share the same value. If we allow the surface of the water to be a distance *h* below the *x*-axis, then *f(x) = -h,* where *h* is a positive number. The function *g(x)* is set equal to $y_-(x)$, which describes the lower section of the ellipse. In order to calculate the integral in (11) explicitly we need to first find the values of X_L and X_R. Figure 4 depicts each of these quantities.

To find the endpoints of the integral, X_L and X_R, we set $-h = y_-(x)$ and solve for *x*. After much simplification, the result is of the form

$$X_{R,L} = \frac{1}{\alpha(1+\beta)}\left[\beta\ h\ \pm\sqrt{ab\alpha\left(1+\beta^2\right)-h^2}\right]. \tag{12}$$

Now that the endpoints have been calculated in terms of *h* and θ, the integral defining the center of mass of the displaced water, (11), can be calculated. The result in terms of the endpoints, X_R and X_L, is as follows

$$\overline{X}_W = \frac{1}{A}\left[\frac{\alpha\beta}{3}\left(X_R{}^3 - X_L{}^3\right) - \frac{h}{2}\left(X_R{}^2 - X_L{}^2\right) - \frac{\alpha}{3}\left[\left(\frac{ab}{\alpha} - X_R{}^2\right)^{\frac{3}{2}} - \left(\frac{ab}{\alpha} - X_L{}^2\right)^{\frac{3}{2}}\right]\right] \qquad (13)$$

Before our expression for $\overline{X}_W$ is complete we must determine *h*. Notice that although the area of the displaced water is a constant, the water level, *h*, is dependent on the orientation of the ellipse. The process of arriving at an expression for *h(θ)* involves calculating the displaced water's area as an integral of the area between two curves. More specifically, the area of the displaced water can be calculated in two ways. The first is as we described above by setting $F_g = F_B$. This method gives us the area of the displaced water. The second method involves computing the integral of the water surface line minus the integral of the elliptical curve (see Figure 4). We can then set these two areas equal to one another and solve for *h*.

The limits of integration are again the *x*-coordinates where the two curves cross, X_L and X_R. The integral determining the area is

$$A = \int_{XL}^{XR}\left[f(x) - g(x)\right]dx \qquad (14)$$

where again $f(x) = -h$, and the function $g(x)$ is set equal to $y_{-}(x)$. This integral yields the result

$$A = \frac{\alpha\beta}{2}\left(X_R{}^2 - X_L{}^2\right) - h\left(X_R - X_L\right) + \frac{\alpha}{2}\left[X_R\sqrt{\frac{ab}{\alpha} - X_R{}^2} - X_L\sqrt{\frac{ab}{\alpha} - X_L{}^2}\right] + \frac{ab}{2}\left[\sin^{-1}\left(X_R\sqrt{\frac{\alpha}{ab}}\right) - \sin^{-1}\left(X_L\sqrt{\frac{\alpha}{ab}}\right)\right] \tag{15}$$

This relation is important, as already mentioned, in that it relates h to the constant area of the water displaced and is a function of the angle θ. Notice that it is a transcendental equation and cannot be solved analytically for h as h is contained in X_R and X_L.

In order for the net torque acting on the system to be zero, and thus, the kayak to be in a state of rotational equilibrium, we must set the torque due to the displaced water acting at $\overline{X}_W$ to be equal to the torque due to the kayak/kayaker system. The force due to the weight of the kayak/kayaker system acts at the center of mass of the system and has an x-coordinate $\overline{X}_K$ given by

$$\overline{X}_K = L\sin\theta \tag{16}$$

where L is the distance of the center of mass of the kayak/kayaker system measured from the center of the elliptical cross-section. From basic physics, L is given by[2]

$$L = \left(\frac{m_{person}}{m_{kayak} + m_{person}}\right)R \tag{17}$$

where *R* is the center of mass of the kayaker relative to the center of the ellipse. Now, setting $\overline{X}_W = \overline{X}_K$, the resulting equation determining mechanical equilibrium is

$$L\sin\theta = \frac{1}{A}\left[\frac{\alpha\beta}{3}\left(X_R{}^3 - X_L{}^3\right) - \frac{h}{2}\left(X_R{}^2 - X_L{}^2\right) - \frac{\alpha}{3}\left[\left(\frac{ab}{\alpha} - X_R{}^2\right)^{\frac{3}{2}} - \left(\frac{ab}{\alpha} - X_L{}^2\right)^{\frac{3}{2}}\right]\right] \tag{18}$$

Solving the System of Equations

We have a set of coupled algebraic equations, (15) and (18), that determine the two variables, *h* and *θ*. To solve the equations numerically, we approximated all of the system's parameters by analyzing a kayak (*a = .38 m*, *b = .15 cm*, *m = 100 kg*) and rearranged equations (15) and (18) into the form

$$0 = \frac{1}{A}\left[\frac{\alpha\beta}{3}\left(X_R{}^3 - X_L{}^3\right) - \frac{h}{2}\left(X_R{}^2 - X_L{}^2\right) - \frac{\alpha}{3}\left[\left(\frac{ab}{\alpha} - X_R{}^2\right)^{\frac{3}{2}} - \left(\frac{ab}{\alpha} - X_L{}^2\right)^{\frac{3}{2}}\right]\right] - L\sin\theta \tag{19}$$

$$0 = \frac{\alpha\beta}{2}\left(X_R{}^2 - X_L{}^2\right) - h\left(X_R - X_L\right) + \frac{\alpha}{2}\left[X_R\sqrt{\frac{ab}{\alpha} - X_R{}^2} - X_L\sqrt{\frac{ab}{\alpha} - X_L{}^2}\right] + \frac{ab}{2}\left[\sin^{-1}\left(X_R\sqrt{\frac{\alpha}{ab}}\right) - \sin^{-1}\left(X_L\sqrt{\frac{\alpha}{ab}}\right)\right] - A. \tag{20}$$

where α and β are defined in (9) and (10) and X_L and X_R are given by (12). We then input the right-hand side of (19) and (20) into a Microsoft Excel spreadsheet and identified the angles where each input changes sign. The water level, *h,* was then manipulated manually until (19) and (20) were both satisfied for the same

value of θ, hence yielding the numerical solution of the above equations for a given choice of parameters. We then allowed the numerical value for the center of mass of the kayak/kayaker system to vary and repeated the above procedure. Figure 5 shows a Microsoft Excel screenshot of this numerical analysis.

We found a range of numerical values for the critical angle, θ_c, for several values of L and plotted θ_c vs. L to display these numerical data points. Intuitively, as the center of mass of the kayak/kayaker system, L, increases, the critical angle, θ_c, should decrease. The plot displaying the numerical data generated shows exactly this with the functional form demonstrated numerically (see Figure 6).

Conclusion

Realistic modeling of even seemingly simple dynamical systems can become increasingly complicated. In many cases, finding an analytical solution of the motion is not feasible without making several assumptions and approximations. Finding the critical angle, θ_c, where the kayak is in a state of unstable equilibrium is an example of such a problem. Finding an analytical solution to this problem is not possible as our system involves a transcendental equation. However, we did succeed in finding values for the critical angle in terms of the center of mass of the kayak-kayaker system via a numerical method.

References

[1] J. B. Marion and S. T. Thornton. *Classical Dynamics of Particles and Systems, 5th Ed*. 2004, 9.
[2] J. M. Pasachoff and R. Wolfson. *Physics for Scientists and Engineers, 3rd Ed*. 1999, 243.
[3] J. K. Vennard, *Elementary Fluid Mechanics*, 4th Ed., 1966, 46-48

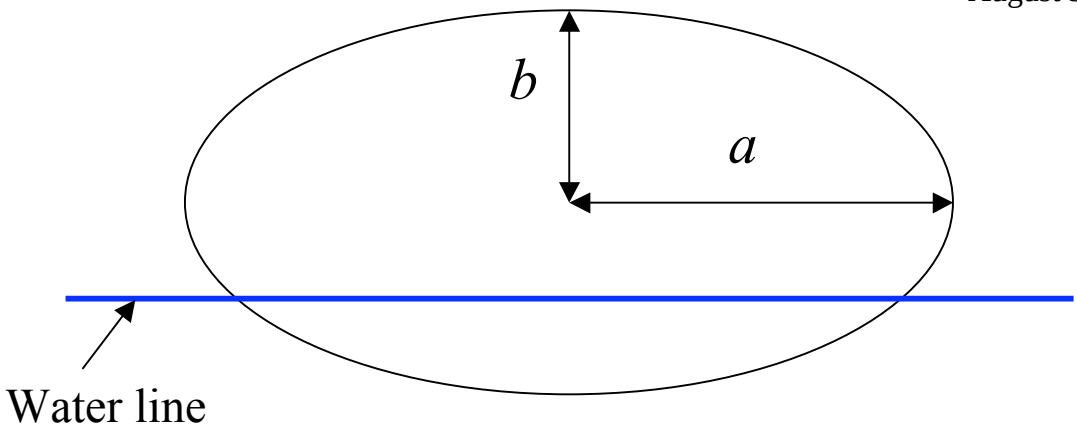


Figure 1. Elliptical cross-section of a kayak

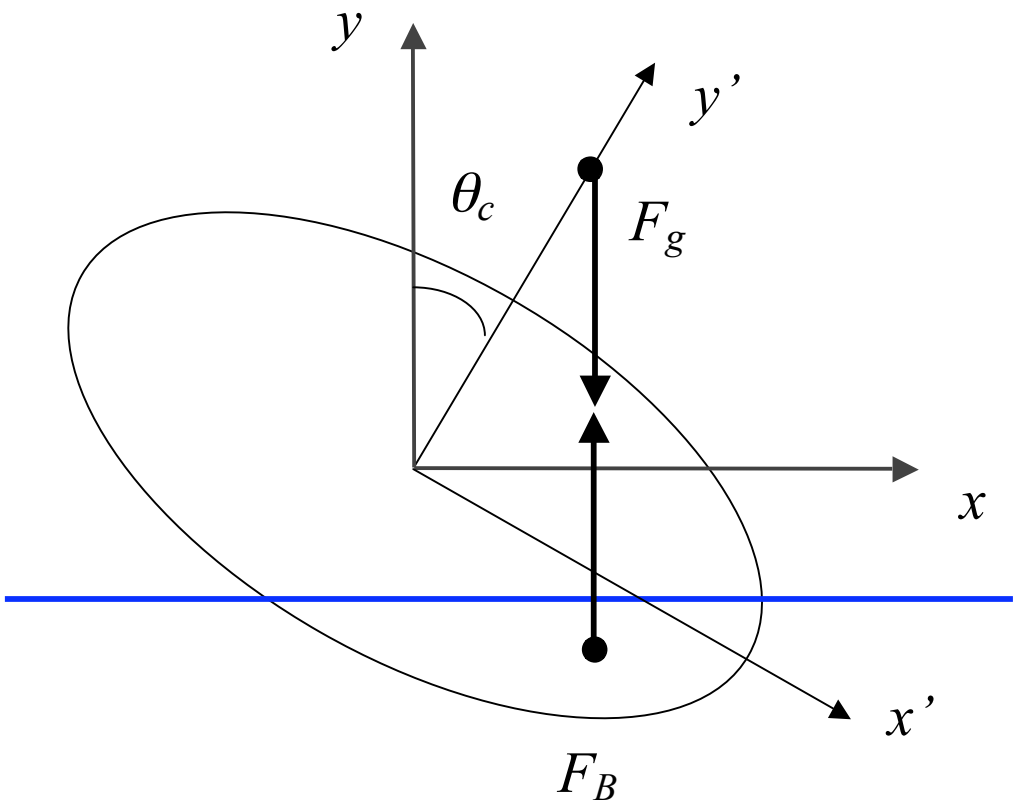


Figure 2. Kayak rotated about an angle

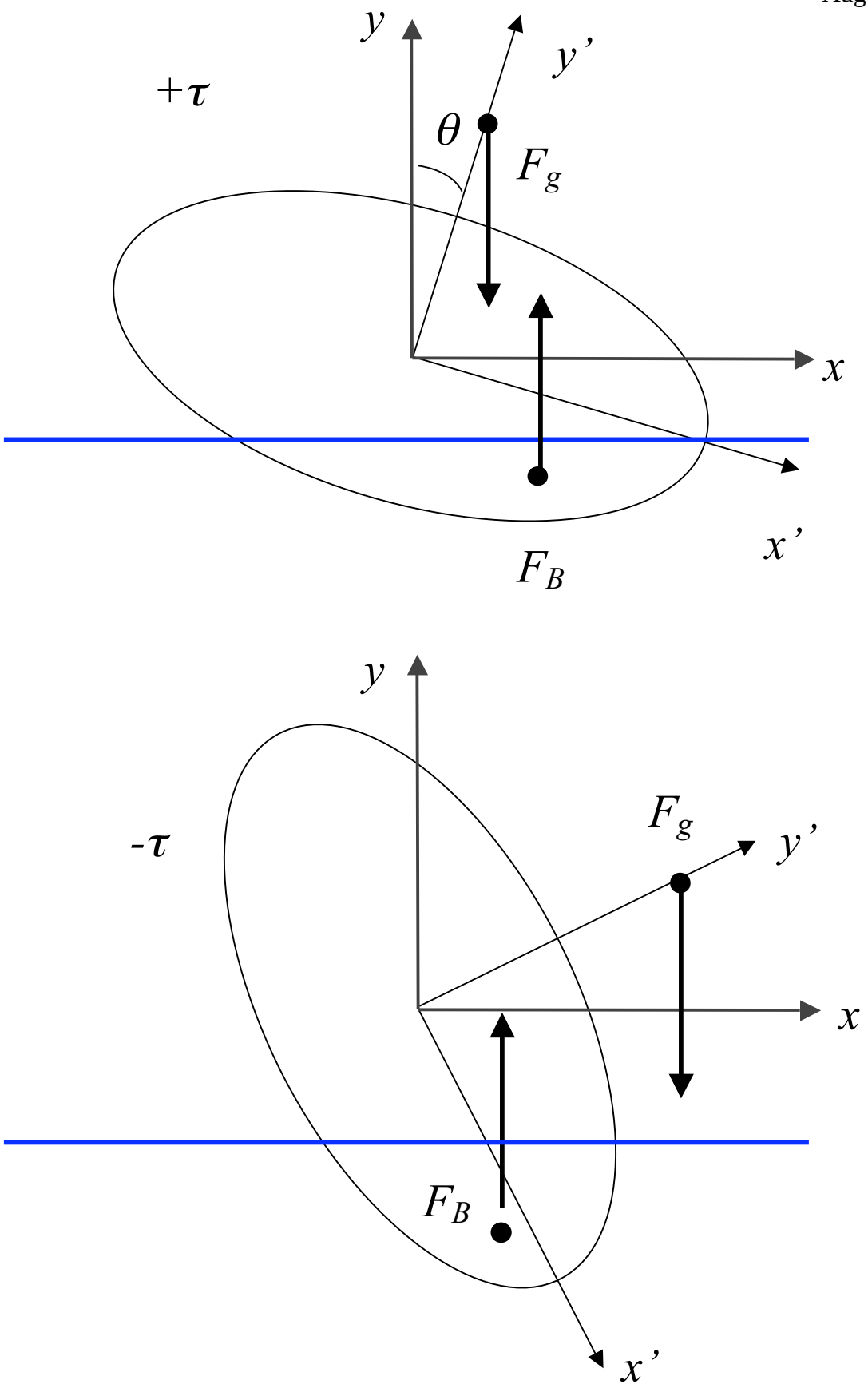


Figure 3. Depiction demonstrating a net torque on a kayak for different angular orientations

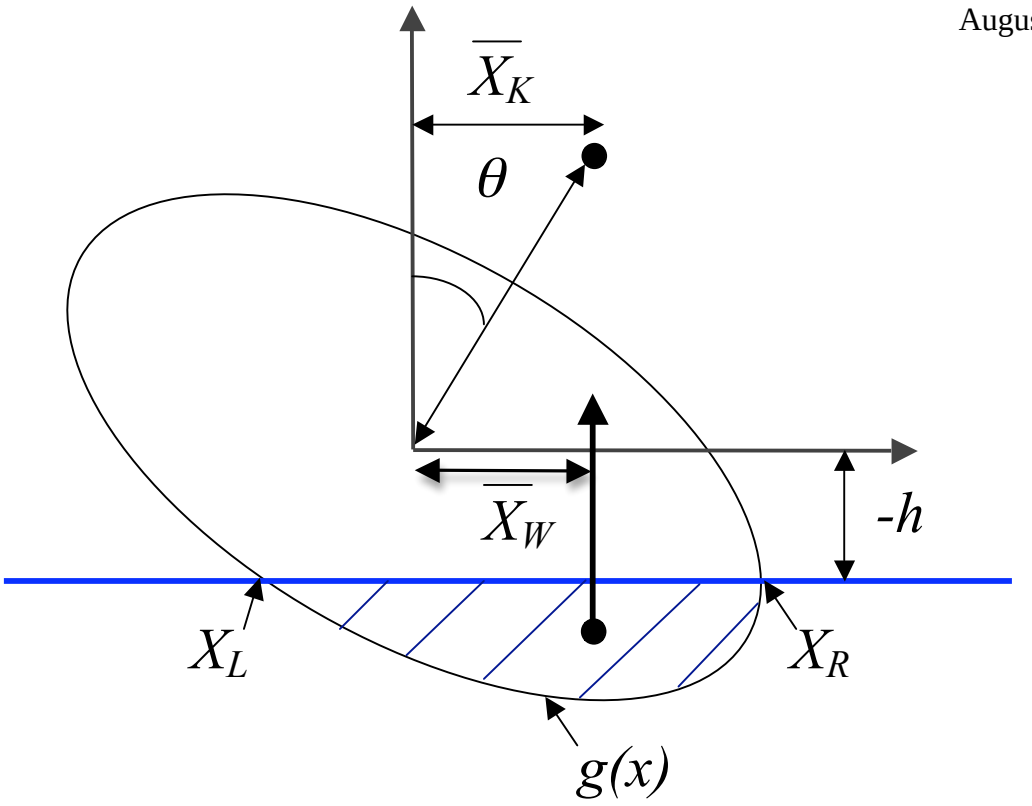
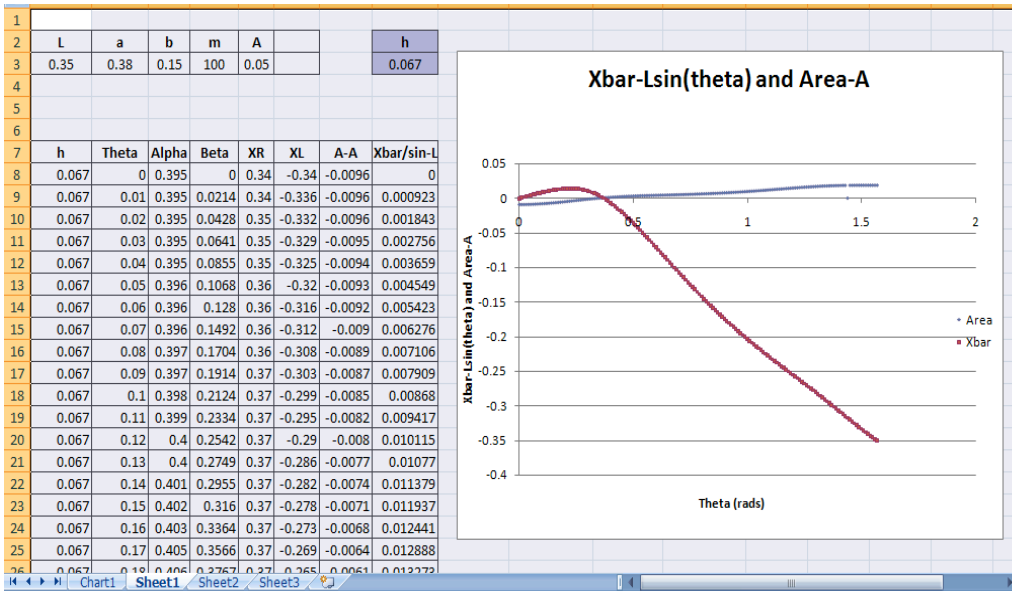


Figure 4. Diagram depicting various aspects of the water displaced by the kayak



L	a	b	m	A			h
0.35	0.38	0.15	100	0.05			0.067

h	Theta	Alpha	Beta	XR	XL	A-A	Xbar/sin-L
0.067	0	0.395	0	0.34	-0.34	-0.0096	0
0.067	0.01	0.395	0.0214	0.34	-0.336	-0.0096	0.000923
0.067	0.02	0.395	0.0428	0.35	-0.332	-0.0096	0.001843
0.067	0.03	0.395	0.0641	0.35	-0.329	-0.0095	0.002756
0.067	0.04	0.395	0.0855	0.35	-0.325	-0.0094	0.003659
0.067	0.05	0.396	0.1068	0.36	-0.32	-0.0093	0.004549
0.067	0.06	0.396	0.128	0.36	-0.316	-0.0092	0.005423
0.067	0.07	0.396	0.1492	0.36	-0.312	-0.009	0.006276
0.067	0.08	0.397	0.1704	0.36	-0.308	-0.0089	0.007106
0.067	0.09	0.397	0.1914	0.37	-0.303	-0.0087	0.007909
0.067	0.1	0.398	0.2124	0.37	-0.299	-0.0085	0.00868
0.067	0.11	0.399	0.2334	0.37	-0.295	-0.0082	0.009417
0.067	0.12	0.4	0.2542	0.37	-0.29	-0.008	0.010115
0.067	0.13	0.4	0.2749	0.37	-0.286	-0.0077	0.01077
0.067	0.14	0.401	0.2955	0.37	-0.282	-0.0074	0.011379
0.067	0.15	0.402	0.316	0.37	-0.278	-0.0071	0.011937
0.067	0.16	0.403	0.3364	0.37	-0.273	-0.0068	0.012441
0.067	0.17	0.405	0.3566	0.37	-0.269	-0.0064	0.012888

Figure 5. Screenshot of numerical method for solving the system of equations

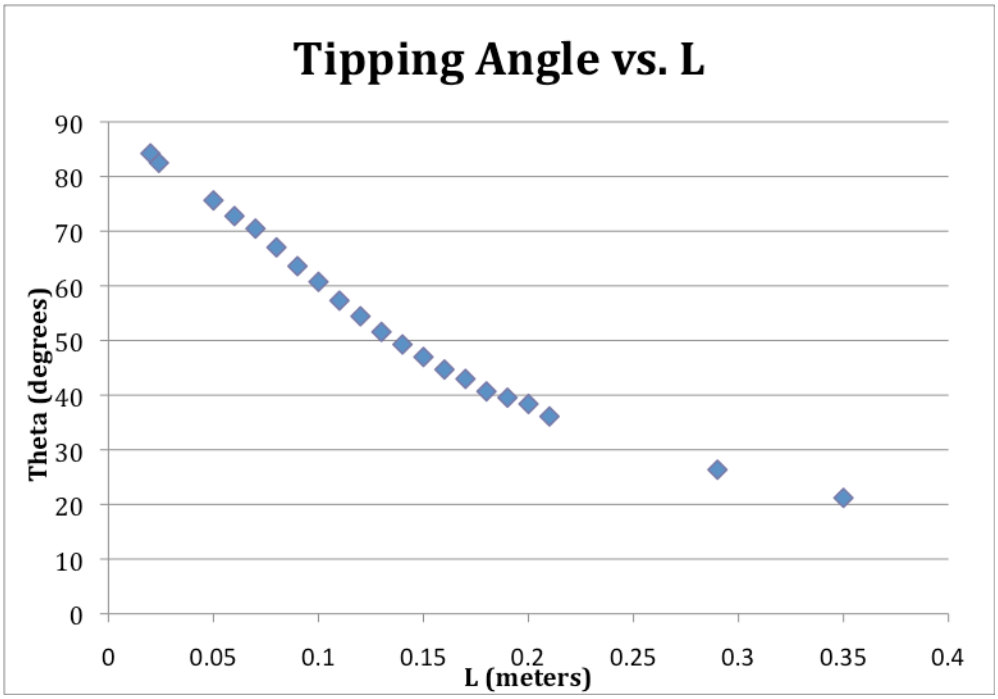


Figure 6. Plot of the critical angle, θ_c, versus the center of mass of the kayak/kayaker system.

Journal of Undergraduate Research in Physics
August 12, 2009

Global Behavior of the Radial Orbit Instablility

Paul A. Lanzel

Eric I. Barnes — Faculty Advisor

Department of Physics, University of Wisconsin — La Crosse, La Crosse, WI 54601

lanzel.paul@ymail.com

barnes.eric@uwlax.edu

ABSTRACT

A variety of evidence points to the existence of dark matter in the universe. As it is not directly observable with conventional astronomical techniques, we must rely on computer models to guide our understanding. We have created a suite of such models in order to observe and explain a specific behavior common to models of dark matter systems called the radial orbit instability (ROI). This instability changes self-gravitating systems from spherical to non-spherical shapes, with corresponding alterations to density and velocity distributions. The initial conditions of our models span a parameter space that is relevant to the radial orbit instability; we control the initial dynamical temperature, density profile, and velocity anisotropy in our models. We have found that in dynamically hot and warm systems, the radial orbit instability will be triggered if the initial velocity anisotropy is high enough. The exact amount of anisotropy required varies somewhat for different initial density profiles. Dynamically cold systems behave somewhat differently, but in general, less initial anisotropy is required to initiate the ROI as compared to hot and warm systems.

1. Introduction

Astronomy and astrophysics are, in some sense, more closely related to archaeology than physics. Like an archaeologist relying on the remnants of eras past, astrophysicists rarely have the luxury of making direct measurements of their subjects and often depend upon inference and indirect means of supporting hypotheses. This is certainly true for those of us who are interested in galaxies. Galaxies are vast collections of stars, gas, and dust of which the NASA Great Observatory space telescopes (Hubble, Chandra, Spitzer, and Compton) continue to provide breathtaking images.

Astrophysicists continue to search for a fundamental picture of how galaxies form and evolve. Since an analogous system of understanding is already largely in place for stars, one might expect that since galaxies are mainly composed of stars, this task would be straightforward. What makes studying galaxies interesting is that the stars are not the most important component of galaxies, at least from the viewpoint of gravity. An analogy to icebergs is often used to clarify our current understanding of galaxies. Stars, gas, and dust (collectively referred to as visible matter since they can produce or absorb light) are the ice that lies above the surface of the ocean. However, given the structure and behavior of this visible ice, we can infer that below the surface is a large, unseen mass. In a galactic context, this mass has become known as dark matter since it does not interact with light in any way. Dark matter neither emits, absorbs, nor reflects light. Given that astronomers mainly use light to detect and study objects in the universe, one is certainly justified in asking, "How can we know there is any dark matter at all?".

The motions of stars and gas in spiral galaxies indicate that individual galaxies are enveloped in dark matter "halos" with roughly 10 to 100 times the total mass of the galaxy's stars (for example, see Salpeter 1978; Rubin 1979; Sancisi & Allen 1979). Additionally, models of the gravitational lensing of light from distant quasars by individual galaxies require similarly large amounts of dark matter to match observations (Schechter & Wambsganss 2004; Ferreras *et al.* 2005). On a very different scale, the motions of individual galaxies within galaxy clusters, which contain hundreds to thousands of galaxies, are also difficult to explain without including large amounts of dark matter throughout the clusters (Zwicky 1937). Recently, detailed observations of the cosmic microwave background (CMB) have been analyzed (Spergel *et al.* 2007). The model of the universe that best explains the distribution of CMB light requires that the bulk of the mass in the universe is not baryonic, or "normal", matter. Specifically, this model universe is composed of matter ($\approx 25\%$ of the total mass-energy budget) and dark energy (the remaining 75% of the budget). While dark energy is a fascinating topic, we will not discuss it further here. Rather, we will focus on the dark matter component, which comprises about 80-90% of the total mass in the universe.

In an effort to understand how dark matter behaves and shapes the formation of galaxies in the universe, astrophysicists have turned to computer models that evolve large numbers of massive particles under the influence of gravity alone. Such models are referred to as N-body models, where N is typically much larger than 10. The large number of particles makes the models approximately collisionless. That is, the motions of the particles are determined by the global distribution of matter in the system, not individual encounters with near-by neighbor particles. Simulations of N-body systems are tools for analyzing the physics behind self-gravitating, collisionless systems of massive particles. We are particularly interested in the physics of models that undergo a collapse phase during their evolution and how collapse shapes a models' long-term behavior.

We have created sets of N-body simulations to investigate a process that is thought to

be important to the formation of dark matter halos, the radial orbit instability (ROI). We provide a brief summary of previous works that deal with the ROI (Polyachenko & Shukhman 1981; van Albada 1982; Merritt & Aguilar 1985; Palmer & Papaloizou 1987; Huss, Jain, & Steinmetz 1999). The ROI is most evident when the particles in an initially spherical system have a high degree of radial motion. Any slight bar-like concentration of particles leads to torques that pull neighboring particles into the bar, increasing the torque strength and perpetuating the instability. The ROI drives such a system towards a prolate spheroidal or triaxial shape. This rearrangement of mass can greatly influence the density distribution and evolution of the halo. Our specific goal is to determine and describe the factors that instigate and drive the ROI. To this end, we focus on models with varying degrees of initial velocity anisotropy; large amounts of radial velocity anisotropy should lead to the ROI, while isotropic velocity distributions should not.

In Section 2, we detail the various initial conditions that influence our simulations and the methods used to evolve our models. Section 3 contains our results as they pertain to the global factors influencing the onset of the ROI. We summarize our work and conclusions in § 4.

2. Methods & Testing

2.1. Initial Conditions

The initial conditions for our models are the control parameters for our investigation. Density profiles, velocity distributions, system mass, and system size are all determined before any simulation begins. Our choices for each are designed to span a wide range of the parameter space relevant to the ROI. Each model discussed here is made up of $N = 10^4$ particles, has a total mass $M = 1$ (in dimensionless code units), and an outer spherical edge with radius $R = 1$ (again, in code units).

We have chosen three initial density distributions to investigate. Models are given uniform ($\rho = \rho_0$), cuspy ($\rho \propto r^{-1}$), or Gaussian ($\rho \propto e^{-r^2}$) radial density profiles. Particle positions are chosen randomly inside the spherical system so that the appropriate distribution is achieved using a Metropolis-like approach. With the particle positions set, the potential energy W of the model is determined.

To complete the initial conditions, we set the velocity of each particle. We do this in two steps; (1) set a scale speed for the system that will determine each particle's speed and (2) choose directions for each vector. The virial theorem is used to relate the total kinetic energy T of a system to its total potential energy $|W|$. Specifically, any gravitationally bound system in virial equilibrium has $T = |W|/2$. For each model, we choose an initial virial ratio

value Q_0, defined by

$$Q_0 = \frac{2T}{|W|}. \tag{1}$$

In this notation, the model is in virial equilibrium when $Q_0 = 1$. We note that throughout this work, any variable with a subscript '0' denotes an initial value. For a given Q_0 and W, the value of T is set. The speed of each particle is chosen so that the average speed for the entire model is $v_{\rm scale}$, where $T \equiv \frac{1}{2}Mv_{\rm scale}^2$. We evolve models with $Q_0 = 1.0$, 0.5, 0.2, and 0.1 to gauge the impact of this quantity on the ROI. Since models with larger values of Q_0 have more kinetic energy, they are often referred to as dynamically hot, keeping with the kinetic theory definition of temperature. We will refer to models with $Q_0 = 1.0$ as "hot", $Q_0 = 0.5$ and 0.2 as "warm", and $Q_0 = 0.1$ as "cold". Changing Q_0 is a simple way to control the strength of the collapse that will occur during a model's evolution. Cold models should have the strongest collapses as the effect of the system's gravity will be to immediately pull all particles towards the center. Conversely, the particles in hot models feel a gravitational pull towards the center, but their substantial initial momenta generally delay and weaken the collapse (for $Q_0 > 1.0$ at least some particles are unbound).

Choosing the directions of each particle's velocity vector requires some care so that we can accurately control the velocity anistropy in a model. The velocity anisotropy β for a set of particles is defined by,

$$\beta \equiv 1 - \frac{\sigma_{\rm tan}^2}{2\sigma_{\rm r}^2}, \tag{2}$$

where $\sigma_{\rm tan}$ and $\sigma_{\rm r}$ are, respectively, the tangential and radial velocity dispersions for that set of particles. The tangential dispersion has two components, one for polar velocities and one for azimuthal velocities; $\sigma_{\rm tan}^2 = \sigma_\theta^2 + \sigma_\phi^2$. Note that the dispersion here is referring to the quantity $\sigma^2 = \langle v^2 \rangle - \langle v \rangle^2$, where the angle brackets indicate averages. The velocity dispersion is most easily thought of as a measure of the spread of the velocities around their mean for a set of particles (in our work, particles within concentric spherical shells). β may have a constant value throughout an entire system, or it may be a function of position within a system. We will briefly discuss a few examples of constant β before continuing on to anisotropy distributions.

In a model where the velocity distribution is isotropic, $\beta = 0$ throughout the system, and a particle's velocity vector points in a randomly chosen direction. Another way of thinking about isotropic velocities is that $\sigma_{\rm r}^2 = \sigma_\theta^2 = \sigma_\phi^2$; the amount of random motion in every direction is the same. This situation also points to the reason for the factor of 2 in the denominator of Equation 2 : $\beta = 0$ when $\sigma_{\rm tan}^2 = 2\sigma_{\rm r}^2$, as in this case. A model with a completely radially anisotropic velocity distribution, $\beta = 1$ everywhere in the model, is constrained to have velocity vectors that are directed towards or away from the center of the system.

We use a flexible anisotropy profile $\beta(r)$ (Barnes *et al.* 2007) to control the amount of radial velocity anisotropy in our models. Depending on their radial distance, some particles

will be set on orbits that are more radially-oriented than others. This flexible $\beta(r)$ is given by

$$\beta(r) = \frac{1}{2}(\beta_{\rm high} - \beta_{\rm low})[1 + \tanh{(m \log r/r_a)}] + \beta_{\rm low}, \tag{3}$$

where β_{low} is the anisotropy value at the center of the system $r = 0$, β_{high} is the anisotropy value at the edge of the system $r = R = 1$, m controls the transition between these values, and r_a is the anisotropy radius at which $\beta = 0.5$. For the models presented here, $\beta_{low} = 0$ and $\beta_{high} = 1$. We use a slope $m = 7$ to give a relatively localized transition between the low and high β values; smaller m would force the transition to occur over a wider range of r, while larger m would make the transition more step-like. Defining β as a function of r gives us the ability to control the quantity and location of the mass that is initially isotropic or radially anisotropic. In general, the central regions of our models have isotropic velocity distributions with a transition to radially anisotropic distributions near the edges. We use r_a as the main control over the anisotropy of a model; small values of r_a drive more particles to be on radially anisotropic orbits, while larger values force the system to be more isotropic (in terms of velocity). Our models include systems with $r_a = 0.1$ to 1.2 in steps of 0.1.

The general method for assigning velocity vectors to fit a given β profile is as follows. For a given particle, a β value is calculated from Equation 3. This value is transformed into an "opening angle" for the velocity vector. Imagine two cones with apexes at the particle and symmetry axes along the radial direction; one cone opens outward and one cone opens towards the center of the halo. When $\beta = 0$, the angle will be π and the velocity vector will be randomly assigned a direction within 4π steradians; this will give an isotropic system. When $\beta = 1$, the angle will be zero and the velocity vector will be completely radial. For values of β between zero and one, the opening angle varies linearly between these extremes.

At this point, we have specified the positions and velocities of each of our N particles. Note that the important control parameters are the density profile $\rho(r)$, the initial virial ratio Q_0, and the anisotropy radius r_a. We now briefly discuss how the particles are evolved in time and analyzed.

2.2. Evolution & Analysis

Models are evolved using the direct N-body integration code NBODY2; for further details on NBODY2 algorithms, see Aarseth (2001). In the code, the gravitational force between any two particles i and j is given by,

$$\vec{F}_{ij} = -G\frac{m_i m_j}{(r_{ij}^2 + \epsilon^2)^{3/2}}(\vec{r_i} - \vec{r_j}) \tag{4}$$

where $G = 1$ in code units, m_i and m_j are the masses of the particles, ϵ is the "softening" length, and r_{ij} is the distance between particles. Note that just as G is dimensionless in code units, so are distances like ϵ and r_{ij}. Basically, these dimensionless distances represent

measurements with respect to the initial radius of the system. It can be seen in Equation 4 that without the inclusion of the softening term ϵ, the gravitational force between two particles approaches infinity as the distance r between the two particles approaches zero. This would, in effect, represent a collisional system which deviates from the seemingly collisionless nature of dark matter. By including the softening value, we set a maximum gravitational force between any two particles and do not allow it to go to infinity. The particles are then essentially collisionless and yet we must take care that the softening is small enough to retain the force's Newtonian character overall. We have tested the collisional nature of our simulations by varying the values for ϵ from $\epsilon = 5 \times 10^{-4}$ to $\epsilon = 5 \times 10^{-2}$, similar to the range suggested by Power *et al.* (2003). Differences between evolutions having ϵ within this range are minimal. The models presented here use the smallest softening value, $\epsilon = 5 \times 10^{-4}$. As a further test of the collisionality of our models, we have also evolved models with particles of different masses. If collisions play a significant role in the evolution, particles with larger masses should "sink" to the center of the system, as collisions with lighter particles should rob them of kinetic energy. We find no evidence of mass segregation in any of these tests, assuring us that our models avoid undesirable two-body collision effects and accurately simulate Newtonian gravity. Evolutions of models with different numbers of particles ($N = 5 \times 10^3$ and $N = 10^4$) have also been conducted. Again, the differences between the various models are minimal, but we have opted to use $N = 10^4$ for increased resolution. Our choices of ϵ and N allow us to perform the large number of simulations necessary to investigate our rather large parameter space while adhering to the collisionless condition required.

The initial density profile determines the crossing time, t_{cross}, which is used as the time interval for NBODY2. Most systems evolve for about 20 t_{cross}; all systems come to virial equilibrium during their evolutions and most reach mechanical equilibrium as well. Depending on the nature of the collapse, some halos that reach virial equilibrium have approximately 5% of their total mass that continues to expand throughout the evolution. During an evolution, NBODY2 regularly produces data that is collected as a "snapshot". These snapshots are measurements of particle numbers, positions, velocities, energies, collision data and more. These time-sequence measurements are then used in the analysis of system shape, density, velocity dispersion, and anisotropy for the systems.

Systemic axis ratios are calculated throughout an evolution using the moment of inertia tensor of the innermost 95% of the mass. We exclude the outermost 5% as it is sometimes not in mechanical equilibrium. These axis ratios are reported as $\frac{b}{a}$ and $\frac{c}{a}$, where a, b, and c are the lengths of the long, intermediate, and short axes of a system. Initially, $\frac{b}{a}$ and $\frac{c}{a}$ are both very close to 1, as one would expect for a spherical system. There are four system shapes that we differentiate. If $\frac{b}{a} \approx \frac{c}{a} > 0.8$, we consider the system to be spherical. We chose this definition largely because systems fulfilling this criteria look spherical, by eye. We consider systems oblate spheroidal when $\frac{b}{a} \gtrsim 0.8$ and $\frac{c}{a} \lesssim 0.8$. If the axis ratios of a system are $\frac{b}{a} \approx \frac{c}{a} < 0.8$, we refer to it as prolate spheroidal. Finally, triaxial systems are those that

have $\frac{b}{a} \neq \frac{c}{a}$, $\frac{b}{a} < 0.8$, and $\frac{c}{a} < 0.8$. In general, to determine whether or not the ROI has occurred, we need only observe whether or not a system has evolved from a spherical shape to one of the other three. Given that $\frac{c}{a}$ compares the shortest axis to the longest, we use its minimum value during an evolution to gauge the maximum strength of the ROI.

We divide systems of particles into concentric spherical shells, each containing 5% of the total mass (typically 500 particles). This gives us a radial profile of quantities like density and velocity distributions. The inner/outer radii and average densities of the shells are saved. The velocity dispersions and β-values for the particles in each shell are also calculated throughout the evolution. We quantify the velocity anisotropy of our models using the fractions of particles that are on isotropic, tangentially anisotropic, and radially anisotropic orbits. The mass within a shell is considered to be isotropic if $-1.0 < \beta < 0.5$, tangentially anisotropic if $\beta < -1.0$, and radially anisotropic if $\beta > 0.5$. For example, if a given model has 5 shells that are isotropic, 10 shells that are radially anisotropic, and 5 shells that are tangentially anisotropic, we would assign an isotropic mass fraction $\mu_i = 0.25$, a radial mass fraction $\mu_r = 0.50$, and a tangential mass fraction $\mu_t = 0.25$. While μ_r and μ_t can be combined to account for the total anisotropic mass fraction μ_a, where $\mu_i + \mu_a = 1$, we distinguish the two types of anisotropic motion to quantify their behaviors and isolate their impacts during the ROI.

Rather than using r_a, which does not have a simple physical interpretation, we use these mass fractions to quantify the anisotropy of our models. Specifically, we use the initial radial mass fraction $\mu_{r,0}$ as one of our parameter space dimensions (initial density profile and Q_0 are the other two). It should be mentioned that μ_r has a monotonic, one-to-one correlation with the global anisotropy parameter $A \equiv 2T_r/T_t$, where T_r is the kinetic energy of radial motions and T_t is the kinetic energy in tangential motions. This global quantity has been used as a predictor of ROI by Polyachenko & Shukhman (1981); Merritt & Aguilar (1985); Barnes *et al.* (1986); Bellovary *et al.* (2008). The benefit of using $\mu_{r,0}$ rather than A_0 is that $\mu_{r,0}$ has a more straightforward interpretation. We point out that the work presented here is a much more thorough examination of parameter space than was attempted in any previous work. We report on the correlations between our parameter space dimensions in the next section.

3. Results

We have found that both Q_0 and $\mu_{r,0}$ influence the onset and evolution of the ROI. First, we look at how the ROI, in hot and warm systems, is affected by increasing $\mu_{r,0}$. Second, we find that cold systems have an initial density profile dependence which influences the ROI. The axis ratios that are discussed in this section are the values of $\frac{b}{a}$ and $\frac{c}{a}$ when $\frac{c}{a}$ is at its minimum for the evolution; *i.e.,*, the maximum departure from spherical symmetry.

3.1. Hot & Warm Models

For hot and warm systems ($Q_0 \geq 0.2$), we find that systems with small $\mu_{r,0}$ retain spherical shapes, while larger values of $\mu_{r,0}$ lead to prolate and triaxial systems. This behavior is independent of initial density profile. Our results show that there is a threshold value for $\mu_{r,0}$ beyond which our models no longer remain spherical; they will evolve into prolate spheroidal or triaxial shapes.

When comparing systems with different initial density profiles, there are three different $\mu_{r,0}$ threshold values that are largely independent of Q_0. An example of this behavior is shown in Figure 1 for a model with $Q_0 = 0.5$ and an initially Gaussian density profile (we

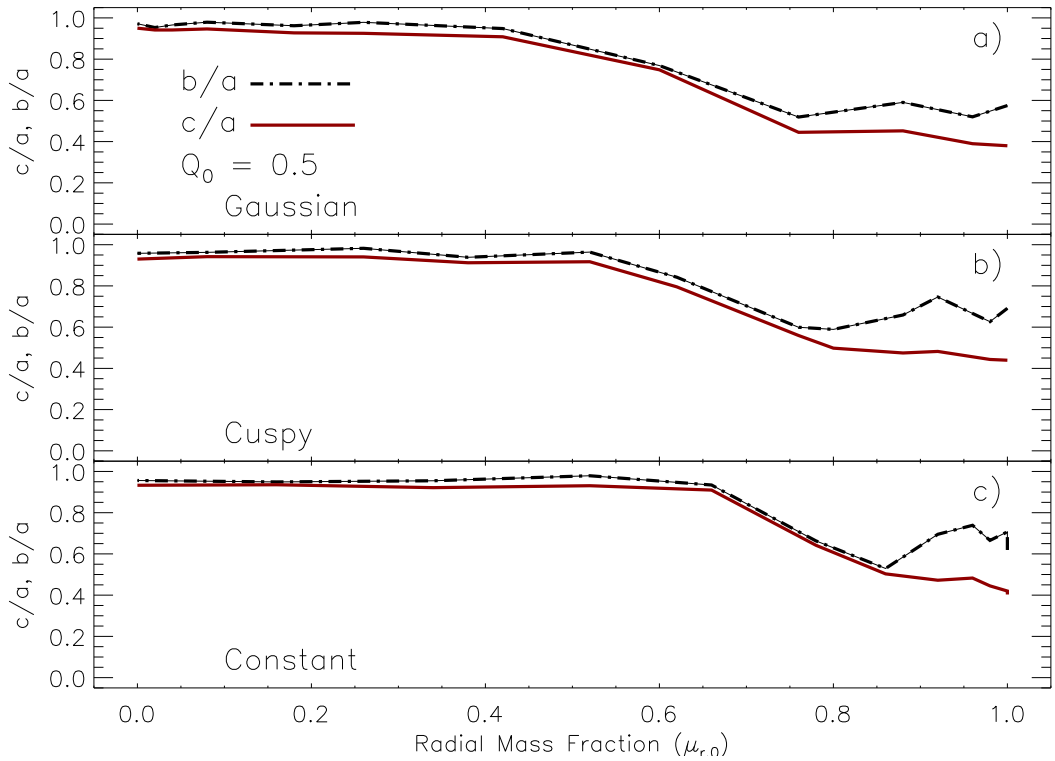


Fig. 1.— Panels a, b, and c show the minimum axis ratios for the innermost 95% of mass as functions of $\mu_{r,0}$ for models with constant, cuspy, and Gaussian initial density profiles, respectively. The dot-dash line represents the minimum $\frac{b}{a}$ axis ratio and the solid line represents the minimum $\frac{c}{a}$ axis ratio. While these specific correlations derive from models with $Q_0 = 0.5$, we find nearly identical behaviors in all models with $Q_0 \geq 0.2$. The spherical-to-non-spherical anisotropy threshold values for the three initial density cases are; $\mu_{r,0} = 0.42$ for Gaussian models, $\mu_{r,0} = 0.52$ for cuspy models, and $\mu_{r,0} = 0.66$ for constant density models.

note that this curve is nearly identical for the $Q_0 = 1.0$ and 0.2 models). When $\mu_{r,0} \leq 0.42$, Gaussian models remain approximately spherical. As $\mu_{r,0}$ increases ($0.42 \leq \mu_{r,0} \leq 0.76$), Gaussian models become prolate spheroidal at the height of the ROI. For models with $\mu_{r,0} \geq 0.76$, the systems become more triaxial as $\mu_{r,0}$ increases; the minimum $\frac{c}{a}$value continues to decline but the minimum $\frac{b}{a}$ value begins to rise. We see a very similar pattern for all three initial density profiles when $Q_0 \geq 0.2$. Initially cuspy and constant density models have spherical-to-non-spherical threshold values of $\mu_{r,0} = 0.52$ and $\mu_{r,0} = 0.66$, respectively.

We explain the observed increase in threshold value as follows. Each density profile (constant, cuspy, and Gaussian) provides a different concentration of particles near the centers of systems. Constant density models will have the fewest particles near the center and Gaussian models will have the most. Any bar-like structure that appears near the center of a system with a high central density can easily ensnare neighboring particles and grow stronger. A similar structure in a system with a relatively lower central density can only grow stronger if particles from the outer regions of the system come into the center where they can be trapped by the growing bar. Because of the imposed shape of the initial anisotropy profile, larger values of $\mu_{r,0}$ correspond to systems with smaller isotropic central regions and larger regions of radially anisotropic mass extending inward from the edge. Systems with lower central density can fall prey to the ROI only if there are enough outer region particles coming into the center, and that requires a larger $\mu_{r,0}$.

3.2. Cold Models

Cold models ($Q_0 = 0.1$) behave somewhat differently than their warm and hot counterparts. The combination of small initial particle velocities and the gravitational pull towards the center of the system greatly increases the number of particles with almost purely radial motion. While this collapsing motion does not initially create high velocity dispersions, particles that have passed through the center of mass quickly create the dispersions needed to initiate the ROI; for more details on cold system collapses related to ROI, see Barnes *et al.* (2009). These delayed, but strong dispersions, lead to an overall increase in the strength of the ROI, but the radial mass fraction $\mu_{r,0}$ is no longer the primary instigator of the ROI due to the low initial particle velocities.

Figure 2 illustrates the behavior of $Q_0 = 0.1$ model axis ratios as a function of initial radial mass fraction for each initial density profile. The Gaussian and cuspy models still follow the same basic pattern as those of the warmer models with the exception that their threshold values for $\mu_{r,0}$ are significantly lower; Gaussian and cuspy models have threshold values of $\mu_{r,0} = 0.18$ and $\mu_{r,0} = 0.26$, respectively. Models with constant density no longer follow the same type of pattern. This is not surprising since many of these cold evolutions lead to large amounts of mass loss (up to $\approx 35\%$) for this initial density profile. The systems that the axis ratios are measuring in these cases are really just "core" remnants of the initial

systems. Again, the ROI will not occur if the bar-like structure near the center cannot grow.

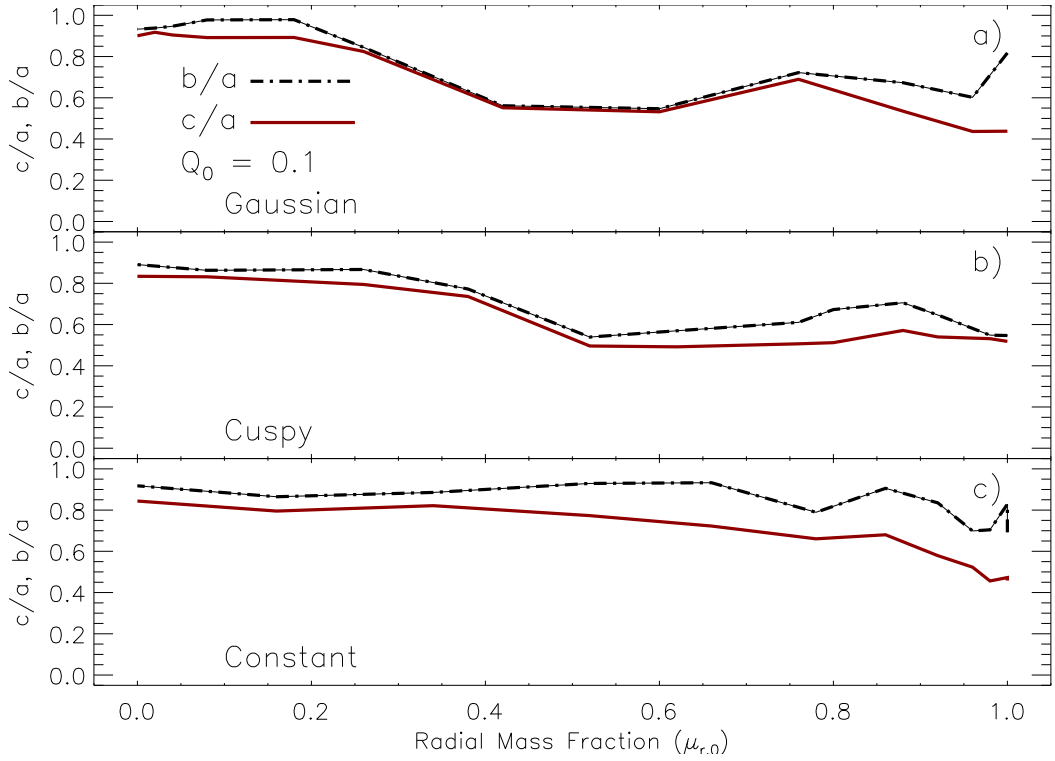


Fig. 2.— As in Figure 1, panels a, b, and c display the minimum axis ratios for the innermost 95% of mass as functions of $\mu_{r,0}$ but for cold models with $Q_0 = 0.1$. From top to bottom, the panels represent models with initially Gaussian, cuspy, and constant density profiles. The dot-dash lines represent the minimum $\frac{b}{a}$ axis ratios and solid lines represent the minimum $\frac{c}{a}$ axis ratios. The threshold values for the Gaussian ($\mu_{r,0} = 0.18$) and cuspy models ($\mu_{r,0} = 0.26$) are substantially lower than in warmer models, but the presence of the threshold remains quite clear. Models with initially constant density profiles do not show clear threshold values. The axis ratios instead have moderately non-spherical values for low $\mu_{r,0}$, and become somewhat more triaxial as $\mu_{r,0}$ increases. We suspect this difference in constant density model behavior stems from the significant mass loss that can occur in these models.

4. Summary & Conclusions

In an effort to increase the understanding of the physics involved in dark matter halo formation and evolution, we have created a suite of models with initial conditions designed to explore a parameter space relevant to the radial orbit instability (ROI). The ROI is an important process in dark matter halo evolution as it changes the overall shape of a system, impacting the density and velocity distributions. These are fundamental quantities that are

necessary when comparing the results of simulations to observations.

We create models with different initial density profiles, initial virial ratios Q_0 and initial radial mass fractions $\mu_{r,0}$. We explore this parameter space by evolving our models with a well-known computer code that accurately follows the behavior of self-gravitating systems. In particular, we are interested in how the overall shapes of our models (quantified by axis ratios) change during their evolutions. Systems that evolve from spherical to either prolate spheroidal or triaxial shapes indicate the presence of the ROI.

We have identified threshold values of $\mu_{r,0}$ that separate systems of particles that evolve spherically from those that become non-spherical. Dynamically hot and warm models have similar behaviors, in that a clear transition is visible in terms of $\mu_{r,0}$ values. Sets of models with each initial density profile show this same behavior, but the specific threshold value changes slightly for each. Models that are dynamically cold react differently than those that are warm. Cold, initially constant density models tend to lose substantial fractions of mass during their evolutions, leaving nearly spherical remnants for almost all values of $\mu_{r,0}$. Initially cuspy and Gaussian models again have threshold $\mu_{r,0}$ values, but the values are much lower than those for the hot and warm models. This is not surprising as the colder initial conditions allow the systems to develop some radial velocity anisotropy through collapse; a smaller amount of initial anisotropy is required to trigger the ROI.

The following chart summarizes how the values for $\mu_{r,0}$ relate to Q_0, initial density profile, and the shape of the system when the ROI is at maximum strength. The smaller $\mu_{r,0}$ number is the threshold value needed to initiate the ROI. Models with smaller values remain spherical. The larger $\mu_{r,0}$ number defines the break between prolate and triaxial systems, with larger $\mu_{r,0}$ leading to triaxial shapes. Systems with intermediate values will result in prolate spheroidal systems. See Figure 3 on the following page for a more visual interpretation.

Q	$\rho(r)$	Prolate Spheroidal Range
1.0	constant	$0.66 < \mu_{r,0} < 0.92$
	cuspy	$0.52 < \mu_{r,0} < 0.80$
	Gaussian	$0.42 < \mu_{r,0} < 0.76$
0.5	constant	$0.66 < \mu_{r,0} < 0.86$
	cuspy	$0.52 < \mu_{r,0} < 0.80$
	Gaussian	$0.42 < \mu_{r,0} < 0.76$
0.2	constant	$0.66 < \mu_{r,0} < 0.86$
	cuspy	$0.52 < \mu_{r,0} < 0.80$
	Gaussian	$0.42 < \mu_{r,0} < 0.88$
0.1	constant	...
	cuspy	$0.26 < \mu_{r,0} < 0.62$
	Gaussian	$0.18 < \mu_{r,0} < 0.76$

Journal of Undergraduate Research in Physics
August 12, 2009

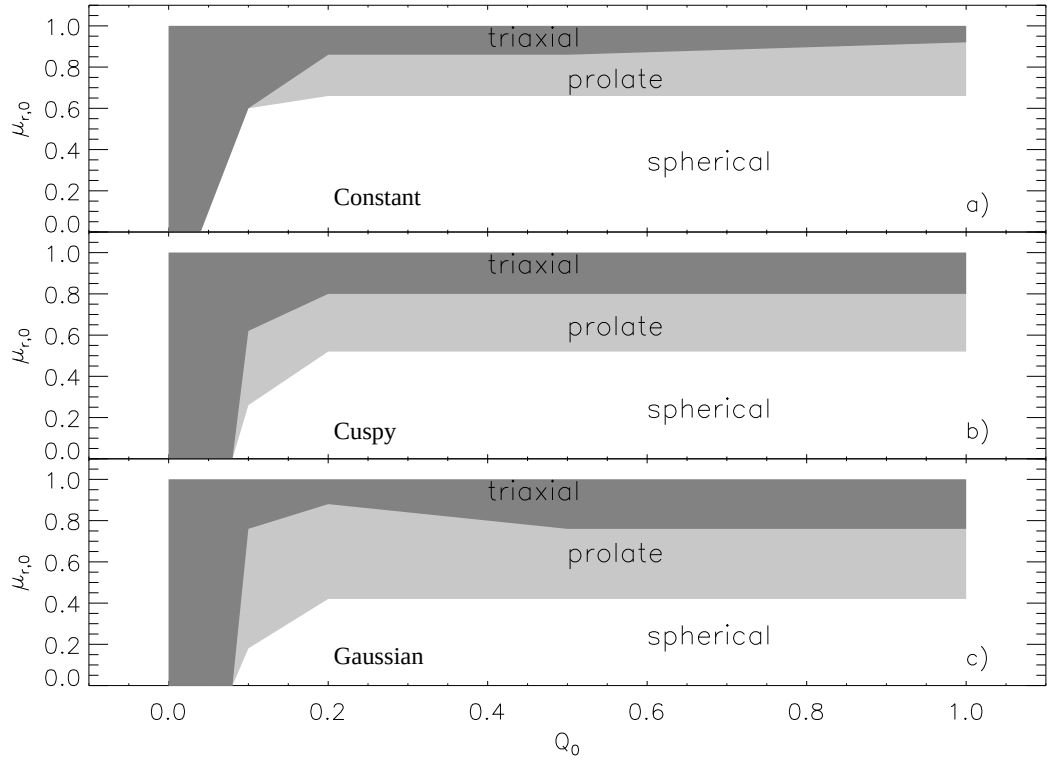


Fig. 3.— The shaded regions in these panels show the boundaries for initial conditions that produce the listed system shapes when the ROI is at its maximum strength. Spherical shapes result from initial conditions in the white areas, prolate shapes result from initial conditions in the light gray areas, and triaxial shapes result from initial conditions in the dark gray areas. Panels a, b, and c represent initially constant, cuspy, and Gaussian density distributions, respectively. In summary, an initially spherical system will evolve into either a prolate or triaxial system if $\mu_{r,0}$ is high enough. While threshold values for $\mu_{r,0}$ are relatively constant for $Q_0 \geq 0.2$, $\mu_{r,0}$ has less of an effect as $Q_0 \lesssim 0.1$.

The authors gratefully acknowledge support from NASA Astrophysics Theory Program grant NNX07AG86G.

REFERENCES

Aarseth, S. J. 2001, New Astron., **6**, 277-291.

Barnes, J., Goodman, J., Hut, P. 1986, Astrophys. J., **300**, 112-131.

Barnes, E. I., Williams, L. L. R., Babul, A., Dalcanton, J. J. 2007, Astrophys. J., **654**, 814-824.

Barnes, E. I., Lanzel, P., Williams, L. L. R. 2009, in preparation.

Bellovary, J. M., Dalcanton, J. J., Babul, A., Quinn, T. R., Maas, R. W., Austin, C. G., Williams, L. L. R., Barnes, E. I. 2008, Astrophys. J., **685**, 739-751.

Ferreras, I., Saha, P., Williams, L. L. R. 2005, Astrophys. J. Lett., **623**, 5-8.

Huss, A., Jain, B., Steinmetz, M. 1999, Astrophys. J., **517**, 64-69.

Merritt, D., Aguilar, L. A. 1985, Mon. Not. Roy. Ast. Soc., **217**, 787-804.

Palmer, P. L., Papaloizou, J. 1987, Mon. Not. Roy. Ast. Soc., **224**, 1043-1053.

Polyachenko, V. L., Shukhman, I. G. 1981, Soviet Astron., **25**, 533.

Rubin, V. C. 1979, Comments Astrophys., **8**, 79-88.

Salpeter, E. E. 1978, in IAU Symp. 77, Structure and Properties of Nearby Galaxies, ed. E. M. Berkhuijsen & R. Wielebinski (Dordrecht:Reidel), 23-26.

Sancisi, R., Allen, R. J. 1979, Astron. & Astrophys., **74**, 73-84.

Schechter, P. L., Wambsganss, J. 2004, in IAU Symp. 220, Dark Matter in Galaxies, ed. S. D. Ryder, D. J. Pisano, M. A. Walker, and K. C. Freeman (San Francisco: Astronomical Society of the Pacific), 103.

Spergel, D. N., Bean, R., Dore, 0., Nolta, M.R., Bennett, C.L., Dunkley, J., Hinshaw, G., Jarosik, N., Komatsu, E., Page, L., Peiris, H.B., Verde, L., Halpern, M., Hill, R.S., Kogut, A., Limon, M., Meyer, S.S., Odegard, N., Tucker, G.S., Weiland, J.L., Wollack, E., Wright, E.L. 2007, Astrophys. J. Supp., **170**, 377-408.

van Albada, T. S. 1982, Mon. Not. Roy. Ast. Soc., **201**, 939-955.

Zwicky, F. 1937, Astrophys. J., **86**, 217.

Key words: Astrophysics, computer simulation, dark matter, Newtonian gravitation, orbit stability, virial theorem

Journal of Undergraduate Research in Physics
March 4, 2009

Time-resolved Photoelectron Spectroscopy and the Photoprotective Properties of Adenine

Amanda N. Brouillette, N. L. Evans, William M. Potter, and Susanne Ullrich

Department of Physics and Astronomy, The University of Georgia, Athens GA 30602

Email undergraduate researcher: anbrou@uga.edu
Email research advisor: ullrich@physast.uga.edu

Abstract

Time-resolved photoelectron spectroscopy (TRPES) is used to measure electronic excited state lifetimes in the DNA base adenine. A detailed description of our femtosecond (fs) laser system, gas-jet molecular beam source, and photoelectron photoion coincidence (PEPICO) spectrometer is given. Ion mass spectra and photoelectron kinetic energy spectra are presented for adenine excitation by 251 nm and ionization by 200 nm. Koopmans'-like ionization correlations are compared to photoelectron spectra, and the states $S_2(\pi\pi^*)$ and $S_1(n\pi^*)$ are identified as participating in the electronic relaxation. We determine that the initially excited $S_2(\pi\pi^*)$ state quickly ($\tau_1 = 71 \pm 16$ fs) decays to populate the $S_1(n\pi^*)$ state, followed by a slow decay to $S_0(\tau_2 = 950 \pm 50$ fs). Our experiments are in good basic agreement with previously reported experiments.[1]

Introduction

Excited electronic states are created in molecules by absorption of UV photons. Molecular dissociation can occur while in this energetically unfavorable excited state. Therefore, relaxation from unstable excited states to the ground state on ultrafast timescales makes some biomolecules stable under UV radiation. Ultrafast relaxation processes have been observed in gas-phase DNA bases and have been the subject of many experimental and theoretical studies due to the inherent significance to the photostability of our genetic material.[2] To this point, adenine has received the most attention, but none of the proposed models of such processes are consistent with all experimental data.

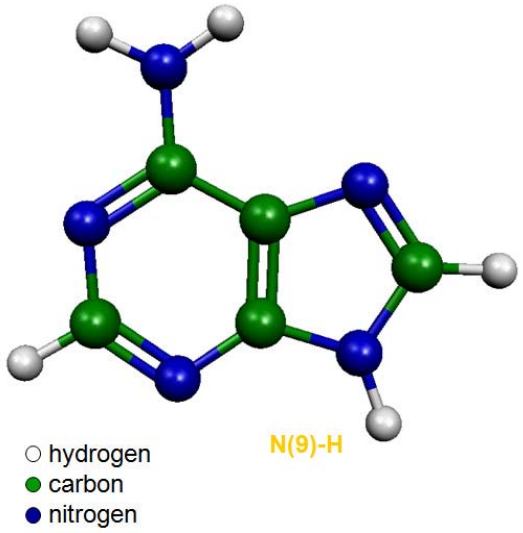


Figure 1. Molecular structure of adenine monomer

Here we focus on the relaxation dynamics of the excited states in adenine which are populated by absorption of UV photons. Only transitions from the ground state, S_0, to low energy excited states are accessible with the UV wavelengths of interest. These transitions involve excitation of electrons in

the highest occupied molecular orbitals (HOMO) into the lowest unoccupied molecular orbitals (LUMO). Specifically, the $S_2(\pi\pi^*)$ state is populated by absorption at 251 nm; subsequent non-radiative decay of the $S_2(\pi\pi^*)$ state creates population in the $S_1(n\pi^*)$ and $S_3(\pi\sigma^*)$ states.

In a very simplistic picture, molecular orbitals (MO) can be thought of as a linear combination of atomic orbitals (AO). In a bonding combination the electron density between nuclei is increased. This causes stabilization whereas an anti-bonding combination reduces the electron density between nuclei and destabilizes the bond. When electrons are excited they obtain an anti-bonding configuration (π^* or σ^*). The photostability of adenine is created by its ability to relax from these anti-bonding orbitals on ultrashort timescales. The atomic character of the shared electrons determine the character of the MO. Shared s-type or in-plane p-type electrons reside in σ orbitals, and shared out-of plane p-type electrons are in π orbitals. Non-bonding electrons such as lone pairs are said to be in n orbitals.

We use the following notation to identify molecular states: S is a singlet state with paired electrons of opposite spin; D is a doublet state with one unpaired electron. A subscript is used to denote the energy level, with S_0 being the ground state and S_1, S_2, and S_3 being the first, second, and third excited state, respectively. We identify the electronically excited state by ($\pi\pi^*$) which denotes promotion of an electron from a π orbital to a π^* orbital. For example, $S_1(n\pi^*)$ is the first excited state, of singlet character, produced by exciting a n electron into a π^* orbital.

Ab initio quantum chemical studies have predicted that several energetically low-lying, singlet excited states play a potential role in adenine's deactivation dynamics. Based on these studies, various relaxation pathways have recently been proposed;[3,4,5] most relevant are the following: Broo's model[4] predicts internal conversion from the initially excited $S_2(\pi\pi^*)$ state to the $S_1(n\pi^*)$ state followed by relaxation back to the S_0 ground state. Puckering of the six-membered ring occurs in the $S_2(\pi\pi^*)$ to $S_1(n\pi^*)$ conversion, and further puckering initiates relaxation back to the S_0 ground state. Sobolewski and Domcke's alternative two-step relaxation pathway[5] involves internal conversion through conical intersections from the $S_2(\pi\pi^*)$ state to the repulsive $S_3(\pi\sigma^*)$ state followed by decay back to the S_0 ground state. Relaxation through the $S_3(\pi\sigma^*)$ state involves elongation of the N(9)-H bond which is located on the five-membered ring as indicated in Figure 1.

Experimentally, the low-lying states of $\pi\pi^*$ and $n\pi^*$ character have been identified in spectrally resolved molecular beam studies[6,7] such as resonance-enhanced multiphoton ionization (REMPI) and laser-induced fluorescence (LIF). The $\pi\sigma^*$ states are difficult to detect spectroscopically as they are optically dark in absorption. Hence, many experiments are limited to indirect probes such as substitution effects and detection of H-atoms released from the N(9)-H group.

Adenine's relaxation dynamics have been studied using time-resolved ion yield measurements. A double exponential decay was observed with time constants of approximately 100 fs and 1.0-1.3 ps following 267 nm excitation.[8,9] Unfortunately, this technique provides no means of directly identifying the excited states involved in the relaxation. The lifetimes changed insignificantly upon deuteration and methyl substitution of the N(9)-H bond. Therefore, it was concluded that the $S_3(\pi\sigma^*)$ state is not involved in the deactivation process.[8] However, this observation is in disagreement with spectroscopic experiments that have detected H-atoms released along the N(9)-H coordinate.[10] Time-resolved photoelectron spectroscopy (TRPES) has recently been used to study the relaxation dynamics in adenine following 267 and 250 nm excitation. These experiments showed a double exponential decay with time constants of 50 fs and 1.2 ps and unambiguously identified the associated $S_2(\pi\pi^*)$ and $S_1(n\pi^*)$ states,

respectively.[1,11] At 267 nm, a decrease in the nπ* amplitude indicates the presence of an additional channel involving the $S_3(\pi\sigma^*)$ state. This assignment has been confirmed through methylation of the N(9)-H group.[11] However, no evidence for the πσ* channel was found in time-resolved photoelectron imaging experiments at this excitation wavelength.[12]

Due to the inconsistencies outlined above we have reinvestigated the photodynamic properties of adenine using TRPES. In this paper, we describe in detail our newly constructed experimental setup and present our first TRPES spectra of adenine recorded at an excitation wavelength of 251 nm. Both time constants and decay-associated spectra reproduce previously reported TRPES results mentioned above.

Experimental Technique

The basis of our TRPES experiment is the familiar concept of the photoelectric effect which states that if light of sufficient energy hits a metal, electrons will be emitted with a very specific amount of kinetic energy. The photoelectron kinetic energy is equal to the energy imparted by incident photon minus the work function of the metal. The work function is the minimum energy required to remove one electron from the surface and is characteristic of the metal. If we consider photoionization of gaseous atoms or molecules, then the work function correlates to the ionization potential (IP) or electron binding energy. Two species with different IP's will therefore produce photoelectrons with different energies. In our application of TRPES we photoionize large biomolecules in the gas-phase and measure photoelectron kinetic energy in order to identify the excited states which participate in ultrafast deactivation pathways.

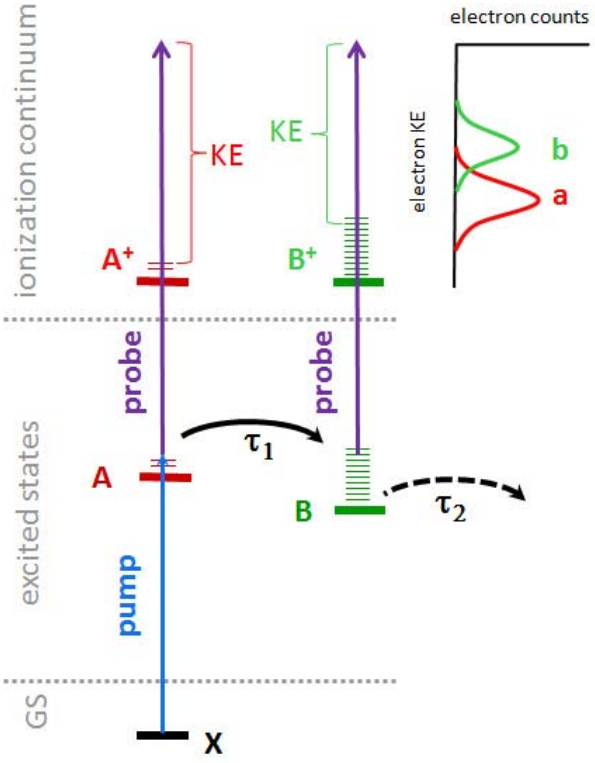


Figure 2. Diagram of a two-state electronic relaxation mechanism.

Deactivation pathways might consist of several internal energy conversion steps, and TRPES provides a unique way to directly identify the electronic character of participating excited states in addition to their lifetimes. The general scheme involves preparation of an excited state, dynamical evolution and a time-delayed probe through ionization. The pump laser populates an electronic excited state by absorption of one photon. The population evolves by internal conversion processes which lead to lower energy electronic levels that carry higher vibrational energy. The time-delayed probe laser then promotes ionization of the population and is thus a measure of the excited states which are present. By changing the time delay between pump and probe pulses one can determine how the population changes with time. In TRPES, photoelectron kinetic energy spectra are measured as a function of the pump-probe delay (Δt) thus providing spectroscopic and dynamic information. To

demonstrate the TRPES experimental technique, a diagram of a two-state relaxation mechanism is shown as Figure 2.

The pump pulse excites a molecule from the ground state to an optically bright excited state *A* which can be coupled to another excited state *B*. State *A* decays into an energetically lower, electronic state *B* in time τ_1. During the relaxation process, the difference in electronic energy is converted into vibrational energy since the molecules are isolated from the environment. The time-delayed probe then promotes ionization of the excited state population. In general, *A* and *B* have different electronic character and preferentially ionize into ionic states *A*+ and *B*+, respectively. Upon ionization, photoelectrons will be emitted with kinetic energy equal to the total photon energy minus the energy of their respective ionic states and the TRPES spectrum will show distinct bands *a* and *b*. Therefore, ionic state energy is determined by measurement of photoelectron kinetic energy. Comparison to He(I) photoelectron spectra provides the assignment for the ionic states and through calculated ionization correlations the excited states can be identified. After excitation, the population in state *A* decreases with time τ_1 and the photoelectron band *a* decays. As *A* decays to populate state *B* the photoelectron band *b* increases. State *B* then decays to the ground state with time τ_2, and the photoelectron band *b* decreases. Measuring the change in these photoelectron bands as a function of Δt reveals the dynamics of each excited state. Using this technique, TRPES is a powerful tool which allows identification of electronic excited states and their associated lifetimes simultaneously.

Experimental Setup

The experimental setup used for these studies consists of three main parts: femtosecond (fs) laser system, gas-jet molecular beam source, and photoelectron photoion coincidence (PEPICO) spectrometer. Details of each component are given here.

Laser System

The commercially available fs laser system from Coherent Inc. consists of Verdi V5 and Evolution 30 pump lasers, a Mira 900 Oscillator, a Legend high-energy amplifier, and an OPerA optical parametric amplifier (OPA). A schematic is shown as Figure 3. A description of IR pulse production and UV pulse conversion follow.

The Mira 900 Ti:Sa oscillator is pumped by the Verdi V5 and produces pulses of 94 ± 7 fs centered on 801 nm with 12 nm bandwidth. This output is split into a seed for the amplifier and a beam for pulse duration measurements using a home-built single-shot autocorrelator.

The Legend high-energy amplifier is based on chirped pulse amplification. The seed pulse is temporally stretched by a diffractive grating, amplified in a Ti:Sa rod, and then temporally recompressed by an opposing grating. This keeps the peak intensity low during amplification and prevents damage to the optical components within the lasing cavity. The Legend output is 2.5 W at 1 kHz repetition rate, centered on 801 nm with 12 nm bandwidth. Amplifier pulse durations of 130 ± 5 fs are measured with a home-built scanning autocorrelator. The output is split to produce both the tunable UV pump pulse and the 200 nm probe pulse.

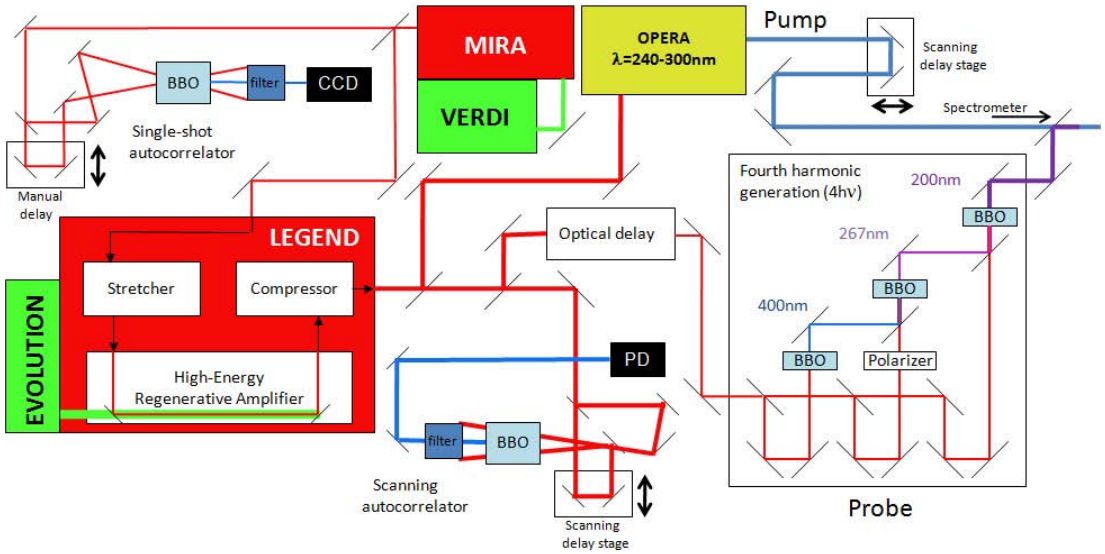


Figure 3. Schematic of Coherent Inc. fs laser system. Also shown are our home-built setup for UV pulse conversion and two autocorrelators.

UV pulses with wavelengths between 240 and 270 nm are created by the Coherent Inc. OPA from 1.0 W of the amplifier output. Inside the OPA, two IR pulses are created and used for conversion to the UV. The conversion is based on sum frequency and second harmonic generation of the IR pulses, and produces 10-20 μJ pulses.

Probe pulses are created from 650 mW of the amplifier output by 4th harmonic generation, using a three stage conversion scheme. A diagram of the conversion is included in Figure 3. In the first step, the 2nd harmonic (400 nm) is produced by frequency doubling in a BBO crystal. The second step produces the 3rd harmonic (267 nm) by mixing the fundamental and 2nd harmonic in BBO. The third step produces the 4th harmonic (200 nm) by mixing the fundamental and 3rd harmonic in BBO. Our home-built conversion setup produces 4-5 μJ probe pulses at 200 nm.

To study the dynamics of molecular excited states we need to observe photoionization events created by one pump photon and one time-delayed probe photon. We must both spatially and temporally overlap our two pulses in order to maximize the desired two-color ionization. Pump and probe beams are combined on a dichroic optic and travel collinearly to the ionization region where they are focused onto the molecular beam by a UVFS lens of 75 cm focal length. Spatial overlap is created by adjusting the pointing of the probe pulse, and temporal overlap is achieved by adjusting the path length of the pump pulse. A motorized delay stage is used to change the pump pulse path length with 0.5 μm precision. The path length (delay position) of the pump pulse which creates the most two-color ionization is termed T_0.

Molecular Beam

A continuous gas-jet molecular beam source prepares adenine molecules in the gas phase. Our molecular beam source chamber is comprised of two differentially pumped vacuum regions and is shown as Figure 4. The basic molecular beam principle is to confine gas at high pressure (HPR), and allow it to expand into vacuum through a micrometer sized pinhole (PH). All but the center portion of the escaping molecules are blocked by two conically shaped apertures (skimmers), which are placed in the beam path between the pinhole and the spectrometer. Each skimmer (S1, S2) has an electroplated opening to reduce turbulence as gas passes, and each of

the skimming stages is differentially pumped to remove the blocked molecules. The molecular beam source provides a high density of molecules in the ionization region while keeping the pressure low in the spectrometer flight tubes.

Gaseous molecular beams are created as described above and are used for instrument calibration, but the adenine we wish to study is in powder form. To create adenine in the gas phase, we heat the sample to 200º C inside the high pressure region. The entire nozzle assembly is temperature controlled by a solid state relay and a resistive band heater (H). Helium gas is supplied continuously to the high pressure region by ¼ in. copper tubing (GS), and carries adenine vapor toward the ionization region upon expansion.

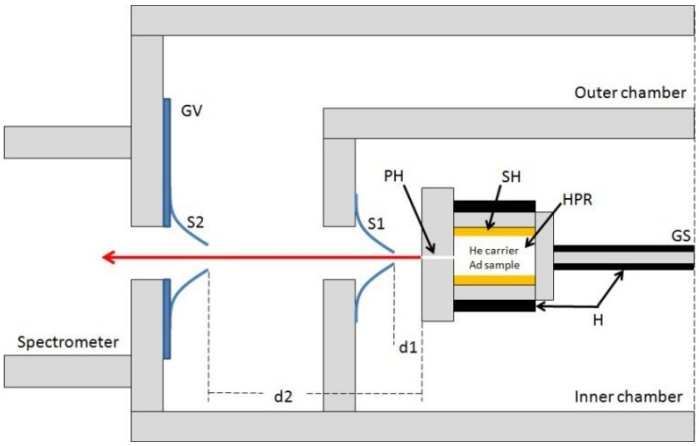


Figure 4. Schematic of molecular beam source chamber. GV-manual gate valve, S1 and S2-skimmer, PH-200μm pinhole, SH-glass sample holder, HPR-high pressure region, GS-gas source, H-heater, d1=15mm, d2=150 mm.

Spectrometer Chamber

A schematic of our PEPICO spectrometer is shown as Figure 5. Pump and probe pulses intersect with the molecular beam in the ionization region (yellow ring) where photoelectrons and positively charged ions are created. Photoelectrons are steered upward by a bottle-shaped magnetic field, and cations are electrostatically accelerated downward where they are detected independently. Photoelectron energy and cation mass are determined by measuring the time of flight (TOF) between photo-ionization and detection. TOF spectra are produced by a multiscaler card, which measures the timing between a reference pulse (start) and a detector pulse (stop). A TOF spectrum is recorded at each pump-probe delay to determine excited state dynamics. Details of the electron and ion spectrometers follow.

Magnetic Bottle Electron Spectrometer

Photoelectrons are emitted with a random angular distribution during ionization. The detector is placed some distance from the ionization region, so electrons emitted away from the detector must be redirected toward the detector. A bottle-shaped magnetic field is used to collect up to 50% of the emitted electrons. [13] This field is created by a strong permanent magnet (RM) in conjunction with a relatively weak solenoid (S). The permanent SmCo magnet is an axially magnetized cylindrical ring that produces an inhomogeneous field of 0.5 T at the maximum along the center axis.

The ionization region (yellow ring) is placed between the solenoid and the ring magnet. The inhomogeneous field directs photoelectrons emitted away from the detector toward the solenoid, where they follow a helical path along the field lines until reaching the electron detector (ED). To negate any electrostatic forces, the ring magnet, entrance grid (EE), ion optics (IO), and electron flight tube are held at ground potential. The energy resolution of the electron spectrometer is determined by the ratio of ring magnet strength to solenoid field strength.[13] The

solenoid is surrounded by a μ-metal magnetic shield (MS), which blocks external fields and allows the solenoid field strength to be kept low.

The photoelectron detector is a 40 mm diameter paired multichannel plate. The detector input is held at ground, where photoelectrons interact with the surface and create secondary electrons within the detector channels. Secondary electrons are multiplied by an avalanche process and attracted to the detector anode, which is held at +4.2kV. The arrival of the secondary electron avalanche at the anode creates a negative pulse of a few ns duration. A capacitive coupling box is used to obtain signals with respect to ground instead of the anode high voltage. Detector pulses are amplified and sent to the multiscaler to produce TOF spectra.

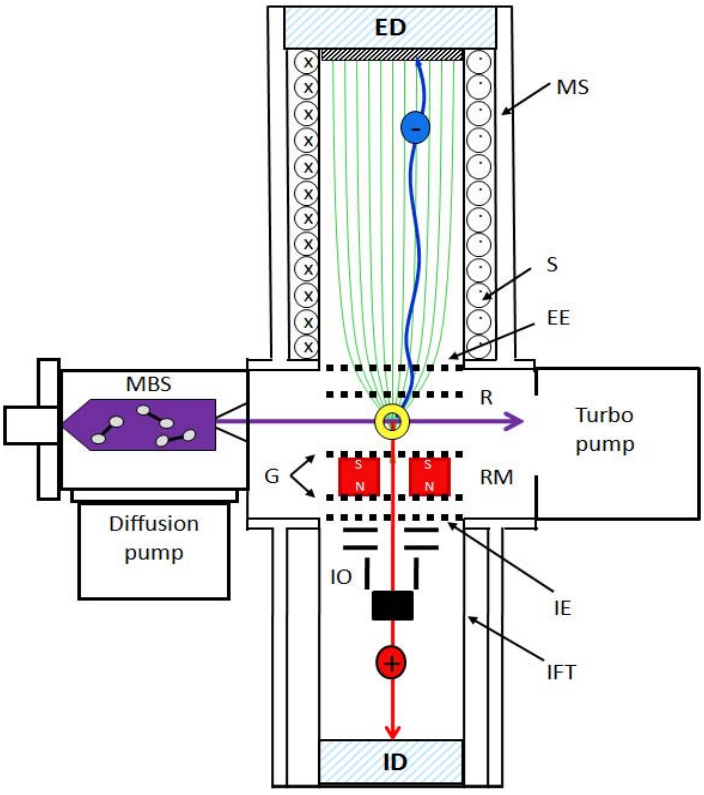


Figure 5. Schematic of PEPICO spectrometer. Laser pulses propagate into the page and the yellow ring marks the ionization region. MBS-molecular beam source, ED-electron detector, MS-magnetic shield, S-solenoid, EE-electron flight tube entrance grid, R-repeller, RM-ring magnet, G-ground, IE-ion flight tube entrance grid, IO-ion optics, IFT-ion flight tube, ID-ion detector.

To determine which excited states of adenine are involved in relaxation, one must know the electron energy, not TOF. 1,3 butadiene is used to convert our TOF spectra into energy spectra. The 1,3 butadiene photoelectron spectrum displays discrete peaks corresponding to different vibrational levels of the ion. The energy of these ionic levels is known from He(I) photoelectron spectroscopy.[14] Assigning TOF peaks to these energy levels serves as calibration.

Ion Spectrometer

Upon ionization, cations have a velocity associated with the molecular beam only, unlike photoelectrons. It is therefore possible to use an electrostatic field to collect all cations created in the ionization region. Ions are accelerated from the ionization region in two steps. First, ions are accelerated by the repeller (R) grid, held at + 1 kV, to ground at the magnet entrance (G). The ring magnet entrance and exit are held at ground, creating an electrostatic field-free region within the magnet. After traversing the magnet, ions are accelerated to the flight tube entrance grid (IE) held at – 1 kV. The flight tube entrance grid and detector input grid are held at the same potential and define the field-free ion flight tube (IFT). Ions travel through the IFT at constant velocities unique to each mass. This creates a temporal separation of ions by mass and allows determination of an ion species by measuring its TOF. Ion optics (IO) inside the drift region correct for the molecular beam velocity. Ions are accelerated from the ion flight tube exit to - 4.2 kV at the ion detector (ID) input. The ion detector is a 20 mm diameter paired multichannel plate

similar to the electron detector. Negative ns pulses are created at the grounded anode, amplified and then sent to the multiscaler in order to produce TOF spectra.

Our TOF ion spectrometer is used to determine the mass of the species present in the molecular beam. TOF peaks produced by ionization of nitric oxide (NO) and 1,3 butadiene are used to convert from TOF to ion mass.

Data Collection

TRPES requires that TOF spectra are recorded while the pump-probe delay is varied. Excited state dynamics are revealed in how the ion mass or photoelectron kinetic energy spectra change as a function of Δt. The entire data collection procedure is computer-controlled, using *LabView*, which controls the delay stage and records TOF spectra according to scan parameters defined before experimentation. These parameters include: starting delay position, delay step size, total delay range, number of times the delay range is traveled and number of start pulses recorded for each TOF spectrum. The delay range is scanned several times to average out long term drifts in laser power or molecular beam intensity, and each TOF spectrum is averaged over many start pulses to account for short term fluctuations. Mechanical shutters placed in the pump and probe beam paths allow measurement of pump- or probe-only (one-color) and pump-probe (two-color) spectra at each delay. Background ionization levels and stable experimental conditions are determined from the one-color spectra. The number of starts to average is controlled independently for the one-color and two-color spectra.

When studying photoelectron spectra from two-photon ionization, background signals arise from two main sources. First is the signal contribution due to multi-photon interactions from each individual wavelength. Second is the contribution to the electron signal from free electrons which are not produced by photoionization of the sample. We aim to measure these background signal levels and subtract them from the total signal in order to study the two-color processes only. The mechanical shutters allow the one-color contribution to be measured in real time. This signal is directly subtracted from the total signal. The free electron background is measured by blocking the molecular beam and recording signal levels from pump and probe pulses separately. We determine that the 200 nm beam produces over 95% of stray electrons in our system. The stray electron contribution is included in the 200 nm one-color signal, and is therefore accounted for in the subtraction step mentioned above.

Results and Discussion

In this section we report decay lifetimes excited states of adenine following excitation at 251 nm. TRPES is used to identify two coupled, excited states which follow the relaxation mechanism described above. Electron energy spectra are used to infer the specific excited state through comparison to Koopmans'-like ionization correlations. The spectra are integrated over electron energy, as described in detail below, and fit with an exponential decay with Gaussian convolution to determine excited state lifetimes.

Our TOF ion spectrometer is used to determine the species present in the molecular beam. The mass spectrum, shown as Figure 7, is produced by co-expansion of heated adenine with 50 torr He carrier gas. Over 98% of ions are from the adenine monomer; therefore the photoelectron spectra we measure are predominantly from adenine.

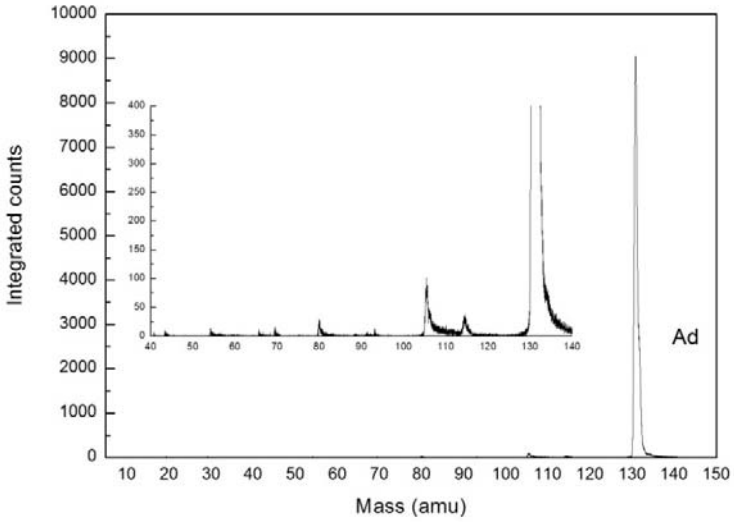


Figure 7. Mass spectrum produced by co-expansion of adenine with 50 torr of He.

Collapsed photoelectron spectra are shown in Figure 8. Matrix A is probe only, matrix B is pump only, and matrix C is pump-probe. Matrices A and B are scaled, due to the difference between the number of start pulses recorded for one-color versus two-color spectra. Matrix D is the background subtracted two-color only signal, produced by subtracting matrices A and B from matrix C. In each plot electron TOF is on the x-axis and pump-probe delay is on the y-axis. Photoelectron counts at each energy-Δt coordinate are represented by the color scale which is the same in each plot. Referring to Figure 2, one can see that it is possible to determine the energy of the ionic states A^+ or B^+ by measuring photoelectron kinetic energy. The ionic state energy is the difference between the total photon energy and the photoelectron kinetic energy, and is termed electron binding energy. Our two-color only TRPES spectrum is converted from TOF to electron binding energy and is shown in the center of Figure 9. This spectrum is used to extract spectroscopic and dynamic information. Visual inspection of this plot shows that the energy spectrum is different in the three different Δt regimes, labeled A, B and C. Different excitation-ionization processes contribute to the two-color signal in each regime. In A the probe is far delayed from the pump, and only the pump-probe process contributes. In B the pump and probe are overlapped, therefore both pump-probe and probe-pump processes contribute. In C the probe comes before the pump, and only the probe-pump process contributes. The electron binding energy spectra are integrated over the three Δt regimes and shown at right of Figure 9. In A the integration is over 1200 fs, in B and C the integration is over 450 fs. These integrated energy spectra are used to identify which excited states participate in the electronic relaxation of adenine.

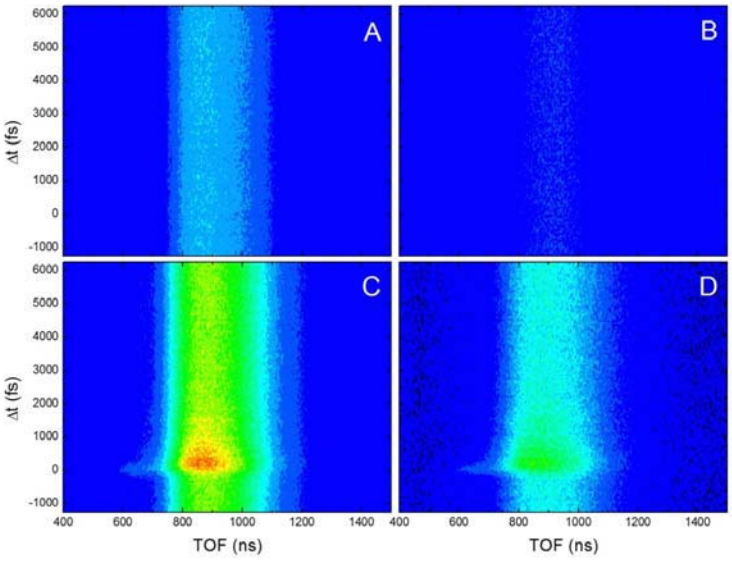


Figure 8. Photoelectron TOF spectra: A-200 nm, B-251 nm, C-200 + 251 nm, D-two-color only.

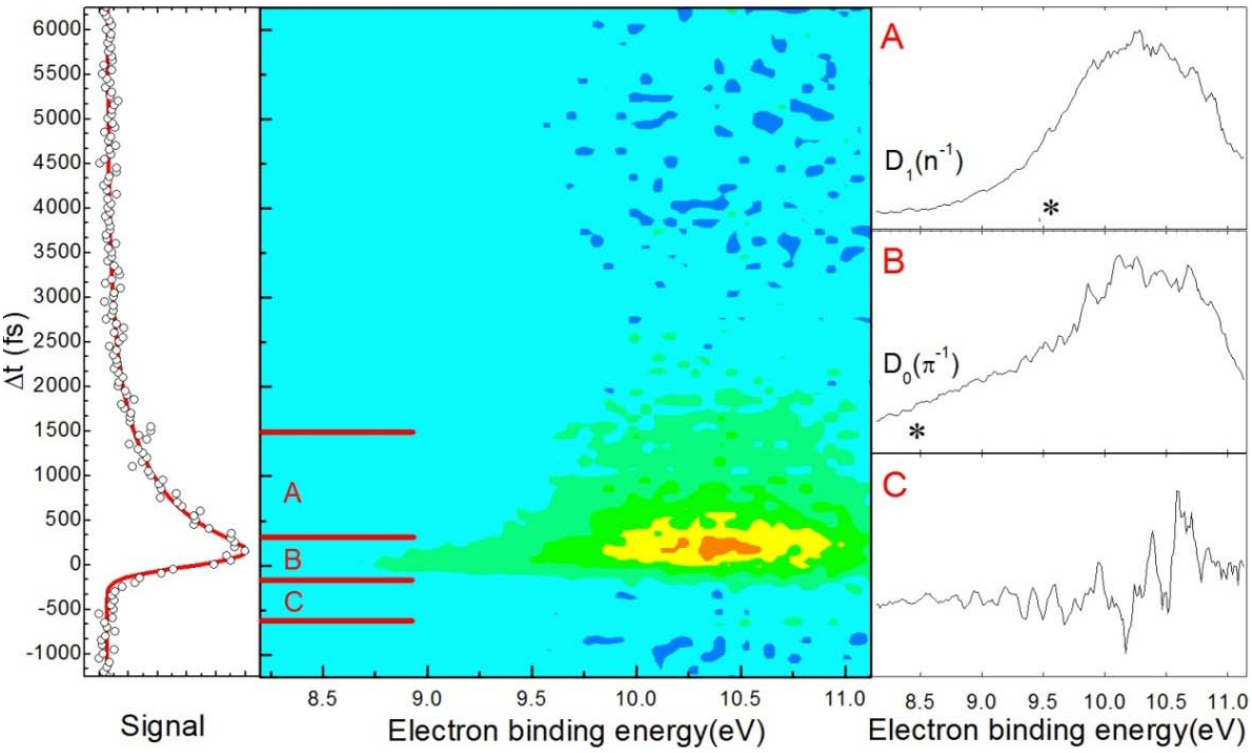


Figure 9. At left is the photoelectron spectra integrated over energy. At center is the electron binding energy spectrum as a function of Δt. At right are the energy spectra integrated over different Δt regimes: A-pump-probe, B-pump-probe and probe-pump, C-probe-pump. Stars show IP for $D_0(\pi^{-1})$ and $D_1(n^{-1})$ cation states.

We assign the bands in our photoelectron spectrum by comparing them to ionization potentials (IP) known from He(I) photoelectron spectroscopy.[15] Two ionization potentials are identified, $IP_0 = 8.5$ eV and $IP_1 = 9.6$ eV, and are shown as stars in plots A and B of Figure 9. IP_0 and IP_1 are from the $D_0(\pi^{-1})$ and $D_1(n^{-1})$ cation states, respectively. In A we assign the pump-probe channel to the $D_1(n^{-1})$ cationic state. In B, a band near 9 eV arises in addition to the pump-probe contribution, and is assigned to the $D_0(\pi^{-1})$ cationic state. The band origin of the neutral excited state (S_2) is near 282 nm,[6] so our energy spectra are shifted by ~ 0.5 eV due to additional vibrational excitation at our pump wavelength. Koopmans'-like ionization correlations have been calculated (TD-B3LYP/6-31++G**) for adenine:[1] S_1, the lowest nπ* state, preferentially ionizes into the $D_1(n^{-1})$ cation excited state, whereas S_2, the lowest ππ* state, and S_3, the lowest πσ* state, both preferentially ionize into the $D_0(\pi^{-1})$ cation ground state. From these ionization correlations we assign the photoelectron band in A to the $S_1(n\pi^*)$ excited state, and the additional photoelectron band in B to $S_2(\pi\pi^*)$.

Lifetimes of the $S_1(n\pi^*)$ and $S_2(\pi\pi^*)$ states are determined independently. For each state, the photoelectron spectrum in Figure 9 is integrated over state specific energy-Δt regions and is fit using the Levenberg-Marquard algorithm. An exponential decay function is convoluted with a Gaussian function with FWHM = 255 fs to account for the temporal duration of our laser pulse. The exponential time constants are determined and identified as the excited state lifetimes.

The photoelectron band for the short-lived $S_2(\pi\pi^*)$ state extends to electron binding energies between 8.75 and 9.5 eV. Several regions in this energy range are integrated over all pump-probe delays, and the lifetime of the $S_2(\pi\pi^*)$ state is determined to be 71 ± 16 fs. Fits are

run with different integration regions to neglect the long-lived state at higher electron binding energies and to determine our error.

The $S_1(n\pi^*)$ state is present in both regions A and B of Figure 9. Due to the overlap with $S_2(\pi\pi^*)$ in region B, only signal at long pump-probe delays is used to find the lifetime of the $S_1(n\pi^*)$ state. The photoelectron spectrum is integrated over all electron binding energies and pump-probe delays greater than 350 fs. The spectrum is fit over several delay regions and the lifetime of the $S_1(n\pi^*)$ state is determined to be 950 ± 50 fs.

Summary

We confirm the following mechanism for adenine relaxation. Excitation by 251 nm is primarily to the bright $S_2(\pi\pi^*)$ state. A fast decay to the $S_1(n\pi^*)$ state occurs in 71 ± 16 fs. The $S_1(n\pi^*)$ state then decays to the ground state with lifetime of 950 ± 50 fs. Future plans include fitting the TRPES data in two dimensions simultaneously which will allow extraction of photoelectron spectra of the $\pi\pi^*$ and $n\pi^*$ states in the overlapping energy-Δt region. Also, the dynamics of adenine at different excitation wavelengths will be studied to identify onsets and branching ratios of competing relaxation pathways, such as the $\pi\sigma^*$ state discussed in the literature.

Acknowledgement

A. N. B. thanks N. L. E. for his mentorship in the laboratory and his contributions to this paper. Acknowledgement is made to the Donors of the American Chemical Society Petroleum Research Fund for partial support of this research (ACS-PRF#44110-G6).

References

1 S. Ullrich, T. Schultz, M. Z. Zgierski, A. Stolow. *J. Am. Chem. Soc., Communicatiom* 2004, **126**, 2262; S. Ullrich, T. Schultz, M. Z. Zgierski, A. Stolow. *Phys. Chem. Chem. Phys.* 2004, **6**, 2796.

2 C. E. Crespo-Hernández, B. Cohen, P. M. Hare, B. Kohler. *Chem. Rev.* 2004, **104**, 1977 and refs. therein.

3 A. L. Sobolewski, W. Domcke, C. Dedonder-Lardeux, C. Jouvet. *Phys. Chem. Chem. Phys.* 2002, **4**, 1093; C. Marian. *J. Chem. Phys.* 2005, **122**, 104314; L. Blancafort. *J. Am. Chem. Soc.* 2006, **128**, 210; H. Chen, S. Li. *J. Phys. Chem.* 2005, **109**, 8443; M. Z. Zgierski, S. Patchkovskii, E. C. Lim. *Can. J. Chem.* 2007, **85**, 124.

4 A. Broo. *J. Phys. Chem. A* 1998, **102**, 526.

5 S. Perun, A. L. Sobolewski, W. Domcke. *Chem. Phys.* 2005, **313**, 107; A. L. Sobolewski, W. Domcke. *Eur. Phys. J. D.* 2002, **20**, 369; S. Perun, A. L. Sobolewski, W. Domcke. *J. Am. Chem. Soc.* 2005, **127**, 6257.

6 N. J. Kim, G. Jeong, Y. S. Kim, J. Sung, S. K. Kim, Y. D. Park. *J. Chem. Phys.* 2000, **113**, 10051.

7 C. Plützer, K. Kleinermanns. *Phys. Chem. Chem. Phys.* 2002, **4**, 4877.

8 H. Kang, B. Jung, S. K. Kim. *J. Chem. Phys.* 2003, **118**, 6717.

9 E. Samoylova, H. Lippert, S. Ullrich, I. V. Hertel, W. Radloff, T. Schultz. *J. Am. Chem. Soc.* 2005, **127**, 1782; C. Canuel, M. Mons, F. Piuzzi, B. Tardivel, I. Dimicoli, M. Elhanine. *J. Chem. Phys.* 2005, **122**, 074316; H. Kang, K. T. Lee, B. Jung, Y. J. Ko, S. K. Kim. *J. Am. Chem. Soc.* 2002, **124**, 12958.

10 K. L. Wells, G. M. Roberts, V. G. Stavros. *Chem. Phys. Lett.* 2007, **446**, 20; I. Hünig, C. Plützer, K. A. Seefeld, D. Lüwenich, M. Nispel, and K. Kleinermanns. *Chem. Phys. Chem.* 2004, **5**, 1427; M. Zierhut, W. Roth, I. Fischer. *Phys. Chem. Chem. Phys.* 2004, **6**, 5178; M. G. D. Nix, A. L. Devine, B. Cronin, M. N. R. Ashfold. *J. Chem. Phys.* 2007, **126**, 124312.

11 H. Satzger, D. Townsend, M. Z. Zgierski, S. Patchkovski, S. Ullrich, A. Stolow. *PNAS* 2006, **103**, 10196.

12 C. Canuel, M. Elhanine, M. Mons, F. Piuzzi, B. Tardivel, I. Dimicoli. *Phys. Chem. Chem. Phys.* 2006, **8**, 3978.

13 P. Kruit, F. H. Read. *J. Phys. E.* 1983, **16**, 313.

14 D. W. Turner, C. Baker, A. D. Baker, C. R. Brundle. *Wiley – Interscience*, 1970.

15 H. Satzger, D. Townsend, A. Stolow. *Chem. Phys. Lett.* 2006, **430**, 144.