CHAOS IN THREE PHYSICAL SYSTEMS

R.S. MACKAY
Mathematics Institute,
University of Warwick,
Coventry CV4 7AL, U.K.
E-mail: mackay@maths.warwick.ac.uk

Examples of various types of chaos are described in some physical systems:

(1) Subshifts of second species orbits in the circular restricted three body problem (with S.V. Bolotin);

(2) Anosov energy levels for the frictionless dynamics of a mechanical linkage, and a normally hyperbolic Anosov submanifold for weak friction and suitable feedback control (with T.J. Hunt);

(3) "Ergodic pumping": a proposed mechanism for the power stroke of myosin, and a design principle for nanobiotechnology (with D.J.C. MacKay).

1. Second species subshifts in the three body problem

The idea that a deterministic system could exhibit chaotic motion arose during Poincaré's work on the three body problem, but he came across it in a near integrable part of the phase space which was hard to analyse because the resulting chaos is exponentially weak in the relevant parameter. In a less well known part of his work[1], Poincaré studied orbits of the three body problem with close encounters. 100 years later we proved that this provides a much more prominent source of chaos[5], as I will review here.

Poincaré's suggestion was that the three body problem of celestial mechanics contains "second species" periodic orbits, i.e. ones which, as the two smaller masses tend to zero, tend to sequences of arcs of Kepler orbits about the heavy mass, joined at collisions between the smaller bodies. It is agreed[2] that he did not have a valid proof and that his claim is false in the generality he stated. Numerical computations[3], however, suggest many such orbits exist, in particular in the planar circular restricted case, where I shall refer to the bodies as the Sun, Jupiter and an Asteroid. Marco and Niederman[4] provided a proof of existence of a second species periodic orbit, more precisely of a one-parameter family with two close encounters of the

Asteroid with Jupiter per period for small enough mass of Jupiter. Bolotin and I[5] extended this dramatically, proving existence of many one-parameter families of chaotic subshifts of second species orbits in the same case.

Theorem 1.1. [5] *In the planar circular restricted three body problem with masses* $1-\varepsilon, \varepsilon, 0$*, for all values* C *in the interval* $(-\sqrt{8}, 3)$ *there are arbitrarily large subsets* T *of rationals and* $\varepsilon_0(T) > 0$ *such that for any sequence* $(m_n/k_n)_{n\in\mathbb{Z}} \in T$ *and* $\varepsilon \in (0, \varepsilon_0)$ *there is a unique trajectory of Jacobi constant* C *near to a chain of collision trajectories formed by transforming to the rotating frame Kepler ellipses of frequencies* m_n/k_n *traversed* m_n *times, and it converges to the chain as* $\varepsilon \to 0$.

The result is hyperbolic horseshoes, which occupy a set of measure zero and are unstable, so their physical relevance is not immediate. Horseshoes like these are, however, the principle of solar system exploration missions like Voyager (though our result is much more restricted because our horseshoes make close encounters with only one planet and assume it is in a circular orbit). The angle change in a close encounter is a smooth function of impact parameter with slope of order $1/\varepsilon$, so these orbits have Lyapunov exponent of the order of $\log 1/\varepsilon$ (in units of the frequency of Jupiter). This is a much stronger instability, and hence offers greater controllability, than schemes using unstable periodic orbits around the colinear Lagrange points, for example, and Sitnikov's horseshoe.

Here is the idea of our proof. In the frame rotating with Jupiter about the Sun, the Jacobi invariant

$$C = -|\dot{q}|^2 + |q|^2 + 2(1-\varepsilon)/|q| + 2\varepsilon/|q-j| - 2\varepsilon x \tag{1}$$

for the Asteroid is preserved, where $q = (x, y)$ is its position and Jupiter is at $j = (1, 0)$. Collisions with Jupiter can be regularised by taking a square root coordinate change about j and rescaling time, to convert the orbits of the Asteroid to those of energy ε of a new C-dependent Hamiltonian which is smooth at j and has a saddle over it at energy 0. The saddle has rotationally symmetric quadratic part. Each segment of Kepler orbit for the Asteroid with $\varepsilon = 0$ that happens to start and end on Jupiter gives a homoclinic orbit to the saddle. In particular, this holds for each rational frequency m/k for which there is no intermediate collision during k revolutions of Jupiter's orbit and m revolutions of the Asteroid's. Figure 1 shows two such ellipses. Each ellipse produces two homoclinic orbits of the regularised system because of the square-root cover. A large subset of the homoclinic orbits are non-degenerate (including all those which make a

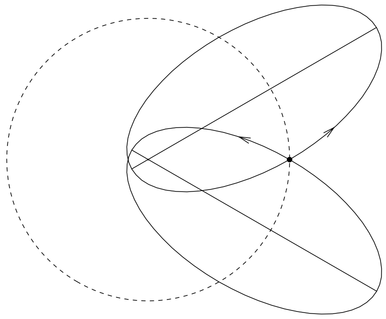

Figure 1. The two orbits of an Asteroid with frequency 1 and given Jacobi constant $C \in (1,3)$, starting and ending at a zero mass Jupiter, in the inertial frame.

complete number of revolutions), i.e. they are non-degenerate critical points of the corresponding Jacobi action functional. Non-degenerate homoclinic orbits to a saddle with rotationally symmetric quadratic part can be glued together to produce nearby true orbits for the regularised system for ε small and positive, if the angle changes are less than $\pi/2$. We proved this by taking a small ball around the saddle and defining a discrete-time action functional for successive entries and exits of the ball using the Jacobi action and then using the implicit function theorem to continue the sequence of collision orbits with respect to ε (this is a perturbation from an "anti-integrable limit"). The resulting orbits of the regularised system convert to true orbits of the Asteroid in the rotating frame if the angle changes of the sequence of Kepler orbits there are never 0 or π (the first keeps the angle changes for the regularised system less than $\pi/2$ if the appropriate branch of the cover is taken, and the second keeps the Asteroid away from collisions with Jupiter).[a]

[a]I take the opportunity to correct an omission from Ref. 5, which does not, however, affect the truth of the main result. In Lemma 3.3 we claimed that for all sequences Ω_n of rationals in a set S_C there is an 'orbite à chocs' consisting of ellipses of frequency Ω_n and Jacobi constant C connecting collisions at Jupiter which leave each collision in neither the same nor the opposite direction as they arrive. We proved this in the inertial frame, but neglected to check the result holds in the rotating frame, which is where we need it to prove our Theorem 2.1. In fact, it holds in the rotating frame for all such sequences except those containing two consecutive Ω_n satisfying $\Omega^{2/3} + 2\Omega = C$. The obstacle in that case is that the two ellipses with such a frequency have angular momentum $+1$, so they transform to orbits in the rotating frame which both leave and arrive at Jupiter tangent to the radial direction. For all other frequencies the transformation of the two ellipses to the rotating frame results in two orbits that leave Jupiter in different non-radial directions, because the angular momentum is $(C - \Omega^{2/3})/2\Omega$.

Some extensions may be possible to the spatial, elliptic, unrestricted and N-body cases.

2. Anosov behaviour in the Triple Linkage

Uniformly hyperbolic attractors, and the *Anosov* case where the whole state space is uniformly hyperbolic, are the ideal type of chaos. There is a very extensive mathematical theory of the consequences of uniform hyperbolicity, which include equivalence to a Markov process and structural stability.

Anosov himself[6], however, has lamented the lack of uniformly hyperbolic attractors in the real world: "the impression is that the Lord prefers to accept a weakened hyperbolicity than labour with the limitations on the topology of the genuine uniformly hyperbolic attractors of the 1960s". Similarly, many authors (e.g. Refs 7, 8) have bemoaned the lack of physical examples of smooth Hamiltonian systems with an Anosov energy level (equivalently an interval of Anosov energy levels). To quote from Hoover[7], "it seems that there are actually *no* interesting smooth differential equations with the Anosov property".

The triple linkage of Thurston and Weeks[9] provides a remedy for both cases[10].[b] It consists of three disks pivoted about points fixed in a roughly equilateral triangle and each connected by a rod from an eccentric point to a common floating pivot (see Figure 2). For a suitable regime of lengths the configuration space is a surface of genus 3, as illustrated in Figure 3.

The free motion (i.e. with no friction nor potential energy) of the linkage is (up to time-scale) the geodesic flow of its *kinetic energy metric*, the metric which assigns to an infinitesimal displacement $v\ dt$ the length $ds = \sqrt{2K(v)}\ dt$, where $K(v)$ is the kinetic energy for velocity v.

[b] There are some almost physical precedents. For example, there is a construction of smooth Hamiltonian systems with an Anosov energy level that we mentioned in Ref. 10 and I am told was already known to Arnol'd: given any Riemannian metric on a surface M of genus greater than 1, there exists a function $V : M \to \mathbb{R}$ such that the Jacobi metric for the Hamiltonian $H(v) = \frac{1}{2}|v|^2 + V(q(v))$ (where $q(v)$ is the base point of v) on the unit energy level of T^*M has negative curvature (even constant) and hence the motion is Anosov. Thus for example, for any length parameters for which the configuration space has a component of genus greater than 1, the free motion of the triple linkage to be described here can be made Anosov for given energy by addition of a suitable potential; but the choice of potential seems to me somewhat contrived. Also, Knauf[11] showed that the regularisation of the motion of a charged particle in a 2D periodic Coulombic lattice is Anosov on some energy levels, but this property does not descend to the original system, and for the principal proposed application – electrons in metals – quantum mechanics, 3D, and electron-electron interactions can not be neglected.

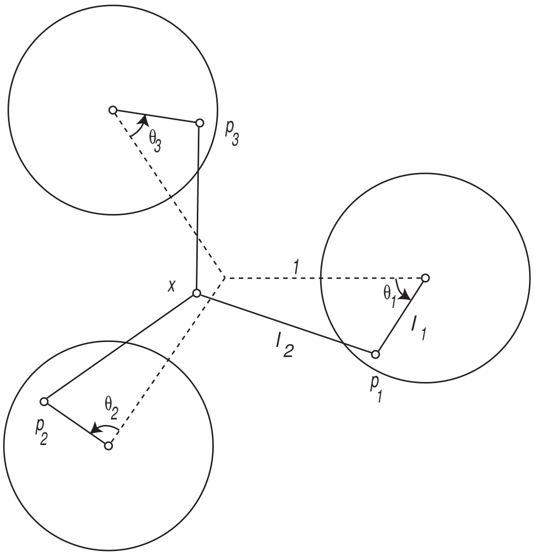

Figure 2. Sketch of the triple linkage.

Theorem 2.1. [10] *For a non-empty open set of lengths and mass distribution, the geodesic flow of the triple linkage is Anosov.*

Here is the idea of the proof. If the system has D_6 symmetry, using the notation of Figure 2, write $\ell_2 = 1 + b\ell_1$ for some parameter b. In the limit as $\ell_1 \to 0$, the configuration space reduces to $\sum_{j=1}^{3} \cos\theta_j = -3b$. In the limit that all the mass is in the disks, the kinetic energy metric is just the restriction of the Euclidean metric $ds = \sqrt{\sum_j d\theta_j^2}$ on $\mathbb{T}^3$ (scaling the moments of inertia I of the disks to 1). Then the curvature of the kinetic energy metric can be computed to be

$$\kappa = \frac{\frac{9}{2}b^2 + 3b\prod_j \cos\theta_j - \frac{1}{2}\sum_j \cos^2\theta_j}{\sum_j \sin^2\theta_j}. \tag{2}$$

For $b = 0$ it is the Schwarz P-surface and has $\kappa \le 0$ with equality iff all $\theta_j \in \{\pm\frac{\pi}{2}\}$. Thus its geodesic flow is Anosov[11]. Anosov systems are structurally stable, so the geodesic flow remains Anosov for all small ℓ_1, b and rod masses, and deviations from D_6-symmetry.

It is worth noting also that all large enough energy levels remain Anosov under addition of a smooth potential, such as provided by gravity if the linkage is not exactly horizontal.

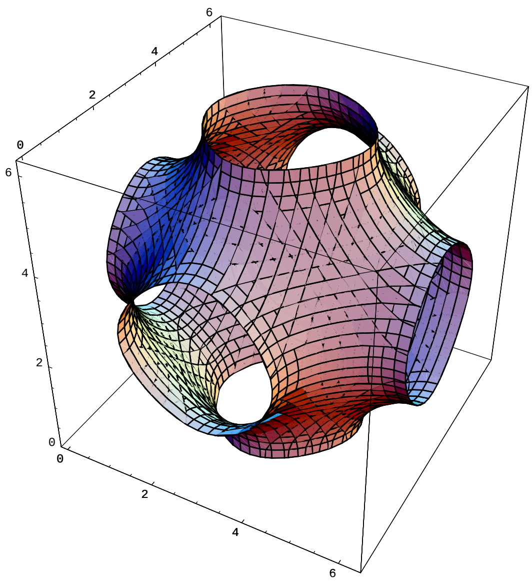

Figure 3. Configuration space for the triple linkage in the space of angles θ_j of the disks, for $l_1 = 19/180, l_2 = 181/180$.

Corollary 2.1. [10] *For Anosov parameter values, the free motion of the triple linkage is mixing, the correlation of any Hölder continuous functions decays exponentially, and the motion in any Abelian cover is diffusive when looked at on a large scale.*

The Engineering department at Warwick made me a triple linkage with parameters which I believe are in the Anosov regime. First impressions confirm that the motion is chaotic, modulo the effects of damping.

Although the frictionless idealisation may satisfy theoretical physicists, it is unlikely to count as physical to an experimentalist, because real physical systems suffer from friction. Nonetheless, the following simple modification produces a system with an attracting submanifold on which the dynamics are Anosov, even in the presence of weak friction.

Choose $E > 0$ (in the case with a potential, choose E sufficiently above its maximum), let $\tilde{K} = \frac{I}{2}\sum_j \dot{\theta}_j^2$ and apply feedback control torques $\Gamma_j = -\gamma(\tilde{K} - E)\dot{\theta}_j$ for some $\gamma > \sqrt{I/2E}$ (or anything C^1-close on a neighbourhood of $\tilde{K}^{-1}(E)$). Then $\tilde{K}^{-1}(E)$ is normally hyperbolic for the limit system and the motion on it is unchanged by the control. Thus for all small smooth perturbations it persists to a normally hyperbolic submanifold carrying Anosov dynamics.

I would be pleased to receive offers to implement such a feedback control.

Note that in principle it should be possible to realise a suspension of a Plykin attractor experimentally, which would be not only uniformly hyperbolic but also "fractal" (a "strange attractor"), but attempts so far produce highly artificial designs[12].

Another good challenge is to demonstrate experimentally some persistent non-uniformly hyperbolic chaotic attractors, like geometric Lorenz attractors (and other persistent singular attractors) and robustly transitive partially hyperbolic attractors (for a summary of results on both, see Sections 6.2 and 6.3 of Ref. 13). Although not structurally stable, they have the property that all C^1-close systems have similar chaotic properties. In particular, persistent singular attractors occur in the unfoldings of various bifurcations, so should be easy to realise experimentally.

3. Ergodic Pumping

While the first two examples use well accepted models for physical systems and exhibit chaos in rigorously proved senses, the third involves a speculative model for a biological system and less precise mathematics.

The question is how free energy decreases can be turned into work in isothermal conditions, with particular reference to myosin, the motor protein in muscle and many other types of cell.

Muscle functions by myosin climbing along actin fibres. The source of free energy for this is hydrolysis of adenosine-triphosphate (ATP), roughly $25T$ per cycle under typical physiological conditions (where T is temperature in energy units). A 50-year challenge (on which much has been written, e.g. Ref. 14 and references therein) is to understand how this free energy is converted to mechanical work.

A large contribution to the free energy change in the power stroke of myosin appears to me to be due to the entropy increase of a phosphate ion (P) and adenosine-diphosphate molecule (ADP) escaping from the ATP binding pocket into the sarcoplasm[c]. I propose that this entropy increase can be converted to work by "ergodic pumping". By this, I mean the time-averaged force exerted on some slow degrees of freedom (Q, P) by some intermediate degrees of freedom (q, p) which are maintained close to thermal equilibrium by many fast ones $(\tilde{q}, \tilde{p})$. In the particular context of

[c]Admittedly, attributing entropy changes in an aqueous environment is delicate, as the organisation of water around biomolecules can have large non-obvious effects.

myosin, the slow degrees of freedom are the large scale conformation of the myosin (in particular the angle of the lever arm), the intermediate ones are the positions of the P, ADP and perhaps an associated magnesium ion (Mg) and proton, and the fast ones are vibrations of the myosin, motion of water molecules, and internal vibrations of the P and ADP.

I first recall how effective stochastic equations arise for slow degrees of freedom, then address the general problem of turning free energy decreases into work, and finally present the mechanism of ergodic pumping and its proposed application to myosin.

A Hamiltonian system with Hamiltonian function H at temperature T has stationary probability measure μ_β whose density with respect to Liouville measure is proportional to $e^{-\beta H}$, where $\beta = 1/T$. Assuming canonical coordinates for simplicity of exposition, the dynamics of the slow degrees of freedom $Z = (Q, P)$ are

$$\dot{Z} = J\nabla H(Z, z), \tag{3}$$

where $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$, ∇ is the gradient with respect to Z, and z stands for the intermediate and fast variables $(q, p, \tilde{q}, \tilde{p})$. Averaging $J\nabla H(Z, z)$ with respect to the conditional measure $\mu_{\beta,Z}$ of μ_β on z given Z splits it into

$$\dot{Z} = J\nabla F_\beta(Z) + \xi(Z, z), \tag{4}$$

where the *free energy* $F_\beta(Z)$ is defined by

$$e^{-\beta F_\beta(Z)} = \int e^{-\beta H(Z,z)}\ dz, \tag{5}$$

with dz denoting the standard volume in $(q, p, \tilde{q}, \tilde{p})$, and $\xi(Z, z)$ has zero mean with respect to $\mu_{\beta,Z}$.

If the Hamiltonian has the simple form $H(Q, P, z) = \frac{1}{2}P^T M^{-1} P + h(Q, z)$ for some positive definite symmetric M then $F_\beta(Q, P)$ takes the form $\frac{1}{2}P^T M^{-1}P + W_\beta(Q)$ for some function W_β (which is commonly called the free energy instead of F_β), so $\dot{Q} = M^{-1}P$ and the Q-component of ξ is zero, but I keep the general setting to allow for more complicated forms of Hamiltonian, which can easily arise. I'll drop the dependence of F on β from the notation from now on.

By Anosov's averaging theorem (e.g. the presentation in Ref. 15), assuming ergodicity of the z dynamics for each fixed Z, one might expect that as the timescale ε for the motion of z goes to 0, the dynamics of Z would converge on any bounded time interval, for all but a set of initial conditions of vanishingly small measure, to the solution of the averaged equations, those where ξ is simply deleted. Unfortunately this is not the relevant scaling limit for most isothermal processes. Instead, the strength of ξ should be considered as large (order $\sqrt{1/\varepsilon}$), but with autocorrelation decaying on a timescale ε, so that the integrated autocorrelation matrix $D_{ij} = \frac{1}{2}\int_{-\infty}^{\infty} \overline{\xi_i(t)\xi_j(0)}\, dt$ converges to a non-zero finite limit, where overbar denotes the average with respect to $\mu_{\beta,Z}$ and the time evolution is the Hamiltonian evolution of z constrained to fixed Z (it is also assumed that $\frac{1}{T}\int_{-T}^{T} |t|\, \overline{\xi_i(t)\xi_j(0)}\, dt \to 0$ as $T \to \infty$).

Convergence of D holds in a variety of circumstances, in particular when the dynamics of the constrained subsystem have suitable forms of chaos, like contact Anosov flows[16] (as for the triple linkage) and certain models with hard collisions (e.g. Sinai's billiard). This is the point at which I use chaos[d]. Indeed, D convergent and positive for generic choices of zero mean observable ξ is quite a good definition of chaos (for the z-motion), and a nice test for this form of chaos has recently been proposed[18]. The matrix D is automatically symmetric and non-negative (in the sense that $u^T D\, u \geq 0$ for all u). $D = 0$ is rare, because roughly speaking that would imply existence of an invariant L^2 graph $Z = \zeta(z)$, and this property is not stable to perturbation of H (see Ref. 17 for some restricted but rigorous results along this line). An interesting project would be to look for an iterative choice of heavy coordinates Z to remove oscillations in the autocorrelation integrand and make its decay as fast as possible, so that accurate estimates of D can be obtained.

Under these assumptions, a remarkable thing happens: even though $\bar{\xi} = 0$, correlations are established between $Z(t)$ and $\xi(Z(t), z(t))$ on a timescale ε so that the mean of $\xi(Z(t), z(t))$ with respect to $\mu_{\beta,Z(0)}$ becomes in general of order 1 for t larger than ε, and produces damping.

The result is a stochastic differential equation in the scaling limit[e], which I think is most simply described by giving its induced evolution on probability densities ρ with respect to volume dZ in Z:

$$\frac{\partial \rho}{\partial t} = -\nabla_i([(J_{ij} - \beta D_{ij})\nabla_j F]\rho - D_{ij}\nabla_j \rho), \text{ for } t > 0, \tag{6}$$

[d] Convergence of D can also often be obtained in the limit of the number of fast degrees of freedom going to infinity, and perhaps this is the more relevant case, as this limit is required in any case to justify the canonical ensemble μ_β.

[e] Although the result is widely believed in the physics community, and the relation between the diffusion and advection exhibited by (6) is known as the "fluctuation-dissipation theorem", I have not yet located a satisfactory proof. There are proofs, however, for some systems in the limit of large number of fast degrees of freedom.[19]

where the summation convention is assumed. This "Fokker-Planck" equation shows the motion to be of advection-diffusion type with diffusion tensor D and advection $\dot{Z}_i = (J_{ij} - \beta D_{ij})\nabla_j F$ (for comparison, the Anosov averaging theorem would give (6) with $D = 0$). Note that (6) has the globally attracting stationary density $\rho_0 = N^{-1}e^{-\beta F}$ (where $N = \int e^{-\beta F(Z)}\, dZ$) in the space of probability densities, which is the marginal of μ_β on Z; furthermore the correction (i.e. the part depending on D) to the probability current (i.e. the argument of ∇_i) is zero for ρ_0 (as it must because $\bar{\xi} = 0$).

A calculation from (6) shows the evolution of the expectation $\langle Z \rangle$ of Z with respect to $\rho(t)$ to be

$$\frac{d}{dt}\langle Z_i \rangle = \langle (J_{ij} - \beta D_{ij})\nabla_j F + \nabla_j D_{ij} \rangle. \tag{7}$$

To show that the correction to the advective velocity produces damping, compute

$$\frac{d}{dt}\langle F(Z) \rangle = \langle -\beta \nabla_i F D_{ij} \nabla_j F + \nabla_i (D_{ij} \nabla_j F) \rangle. \tag{8}$$

Since D is nonnegative, if we keep only the first term we see that $\langle F(Z) \rangle$ decreases as time increases except for densities ρ supported on critical points of F (or places where ∇F is a null vector of D, but typically the term $J\nabla F$ carries Z off all such places except for critical points), so it indeed dissipates free energy. The second term of (8) provides a possible source of increase of $\langle F(Z) \rangle$ when ∇F is small. This represents the fact for example that fluctuations from a local minimum of F push its mean value higher by about T times the numbers of degrees of freedom of Z, to achieve the equilibrium density ρ_0.

The message is that in this scaling limit, F decreases along the path followed by $Z(t)$ except in regions where the dissipation rate $\beta \nabla_i F D_{ij} \nabla_j F$ is less than the diffusive source $\nabla_i (D_{ij} \nabla_j F)$.

Now consider the problem of turning free energy decreases into work. This means we take a Hamiltonian of the form $H_{ext}(Z) + H_{int}(Z, z)$, where H_{ext} is considered as "external energy". Then

$$F(Z) = H_{ext}(Z) + F_{int}(Z), \tag{9}$$

with F_{int} defined by

$$e^{-\beta F_{int}(Z)} = \int e^{-\beta H_{int}(Z,z)}\, dz. \tag{10}$$

Suppose Z is started in a state which is not a minimum of F_{int}. To obtain a significant increase in H_{ext} we had better have at least a corresponding and nearly monotone decrease in F_{int}, and to convert a large fraction of the decrease in F_{int} to increase in H_{ext} we had better have the dissipation rate $\beta \nabla_i F D_{ij} \nabla_j F$ small along the path. An internal free energy decrease can be converted to work with high efficiency if and only if the function H_{ext} matches F_{int} well in the sense that there is a probable path for (6) with only small downhill slope for $F = F_{int} + H_{ext}$.

To demonstrate ergodic pumping, D.J.C. MacKay simulated the following toy model of myosin: the binding pocket is a two-dimensional wedge closed by a piston whose width adjusts to fill the aperture; the piston is subject to a potential of the form $U(Q) = -\frac{A}{Q} + \frac{A'}{6Q^6} + BQ$, where Q is the distance of the piston from the apex of the wedge, which represents a Coulomb attraction of the piston to the apex, a repulsion at short distances and a constant opposing force; it has two drivers behind it (a crude way of taking into account both the P and ADP); the walls are represented by short-range half-quadratic potentials; the dynamics and thermalisation (both at the walls and in the bulk) are simulated by using a Metropolis accept/reject algorithm for leapfrog timesteps of the system of two drivers plus piston, based on the energy change. The free energy W_β for piston position Q is approximately $U(Q) - TS(Q)$ where the *entropy* $S(Q) = 2 \log \alpha Q^2$ with α half the tangent of the angle of the wedge (the entropy comes from the accessible volume αQ^2 for each of the two drivers). Of this, we regard U as external energy. These functions are plotted in Figure 4 for $T = 1$, along with the results of several simulations, showing that from the entropy increase of approximately 24, in expansion from the little well in the free energy at about $Q = 0.1$ to the equilibrium around $Q = 40$, about $16T$ work is done. This is the order of magnitude of work per cycle and efficiency of real muscle (under optimal load/velocity, rabbit skeletal muscle delivers about $15T$ work per ATP hydrolysis, out of a total of $25T$ free energy decrease). Note that the drivers do work just by their entropy increase, without changing energy on average. They achieve this by sucking heat out of the thermal bath simulated by the Metropolis algorithm.

It would be nice to demonstrate ergodic pumping for a fully deterministic model, though it would require many degrees of freedom to keep the temperature close to constant. Suitable ones where convergence of D might be provable could probably be based on Wojtkowski's bouncing balls[20], replacing the floor by a heavy particle in a different potential from the others.

I propose that the power stroke of myosin operates by ergodic pumping. The starting point of the power stroke is myosin with an ADP and P in its ATP binding pocket and just beginning to bind to actin. In this state

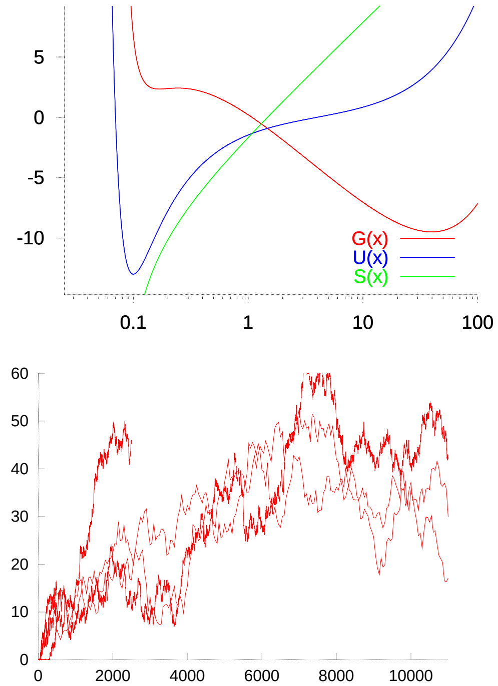

Figure 4. Example of ergodic pumping: (a) Entropy S, external potential U and total free energy $G = U - TS$ as functions of the piston position for the model ($T = 1$); (b) Position of the piston as a function of time for several simulations, each starting at $Q_0 = 0.1$.

the centre of mass of the P has of the order of 1Å^3 available to it (to set the scale, the length of a typical $C - C$ bond in the primary chain of a protein is about 1.5Å), whereas the concentration of P in the cellular fluid is about 2 mM (i.e. 800,000Å^3 per P). So escape of P from the pocket could lead to an entropy increase of about $\log 800,000 \approx 13$ (modulo entropy effects due to changes in its immediate environment, like structuring of water). This could be converted into up to $13T$ of work against an external potential by ergodic pumping. Indeed the "switch II linker, helix and loop" move away from the rest of the pocket and neighbouring parts by about 5Å during the power stroke and this movement is amplified by a relatively rigid "lever arm" to a sliding of the myosin tail relative to the actin fibre by about 100Å, depending on the type of myosin. The "backdoor" of the ATP binding pocket is opened in the process, allowing the P a successively increasing volume. When the power stroke is complete the P becomes free to escape and new ones can enter[f].

Actually, the processes in myosin and muscle are more complicated than this (see Ref. 21 for a summary of the main issues). Nevertheless, I think there is a good chance that ergodic pumping is the main mechanism of the power stroke of myosin. I also speculate[21] that ergodic pumping is a significant aspect of the mechanism of many other biomolecular conformation changes. Furthermore, ergodic pumping could be a good design principle for generation of force and work in nanobiotechnology.

To conclude, the viewpoint I am emphasising is to consider biomolecular conformation changes as continuous processes with potentially non-trivial non-equilibrium dynamics rather than trying to reduce them to the standard chemistry picture of a Markov process of transitions between discrete states (as exemplified by Refs 22, 23). The latter view can be justified if there are fairly high free energy barriers (compared with T) which separate the state space for the heavy variables into identifiable free energy wells with rapid internal equilibration; then the rate of transition is limited by the successful attempts to pass over the barrier, for which there is a well developed large deviation theory. I suspect, however, that the power stroke has only a small free energy barrier at the beginning, if at all, so that virtually every close encounter of the myosin and actin binding sites leads inevitably to a power stroke and the time course for the power stroke is not describable by standard kinetics.

Acknowledgements

I am grateful to David MacKay for carrying out the simulations described in Section 3 and contributing to the scientific development of the idea, and for detailed comments on the presentation of the whole manuscript.

References

1. Poincaré H, Les Méthodes Nouvelles de la Mécanique Céleste, tome III, Gauthiers-Villars, 1899, Ch XXXII.
2. Levy P, Editorial notes, p.629 of Poincare H, Oeuvres tome VII, Gauthiers-Villars, 1952.
3. Hénon MR, Generating families in the restricted three body problem, Lect Notes Phys Monographs 52 (Springer, 1997).

[f] We didn't include this effect in our model, but it would be relatively easy by replacing the hard collisions with the piston by a potential wall whose height decreases as Q increases, so that once Q reaches 40 the particles have a large chance of crossing the piston.

4. Marco J-P, Niederman L, Sur la construction des solutions de seconde espèce dans le problème plan restreint des trois corps, Ann Inst H Poincaré Phys Théor 62 (1995) 211–249.
5. Bolotin SV, MacKay RS, Periodic and chaotic trajectories of the second species for the n-centre problem, Cel Mech Dyn Astron 77 (2000) 49–75.
6. Interview with D.V. Anosov (part 2), Eur Math Soc Newsletter 48 (June 2003) 15–20.
7. Hoover WG, Time reversibility, computer simulation, and chaos (World Sci, 1999), p.219
8. Szydlowski M, Curvature of gravitationally bound mechanical systems, J Math Phys 35 (1994) 1850–80.
9. Thurston WP, Weeks JR, The mathematics of three-dimensional manifolds, Sci Am 251 (1984) 94–106.
10. Hunt TJ, MacKay RS, Anosov parameter values for the triple linkage and a physical system with a uniformly chaotic attractor, Nonlinearity 16 (2003) 1499-1510.
11. Knauf A, Ergodic and topological properties of Coulombic periodic potentials, Commun Math Phys 110 (1987) 89–112.
12. Hunt TJ, Low dimensional dynamics: bifurcations of cantori and realisations of uniform hyperbolicity, PhD thesis, DAMTP, Cambridge (2000).
13. Viana M, Stochastic dynamics of deterministic systems (IMPA, 1997).
14. Howard J, Mechanics of motor proteins and the cytoskeleton (Sinauer, 2001).
15. Lochak P, Meunier C, Multiphase averaging for classical systems (Springer, 1988).
16. Liverani C, On contact Anosov flows, arXiv.math.DS/0303237
17. Field M, Melbourne I, Török A, Stability of mixing for hyperbolic flows, preprint.
18. Gottwald GA, Melbourne I, A new test for chaos in deterministic systems, subm to Proc Roy Soc Lond A.
19. Givon D, Kupferman R, Stuart A, Extracting macroscopic dynamics: model problems and algorithms, Nonlinearity, to appear.
20. Wojtkowski MP, The system of one dimensional balls in an external field II, Commun Math Phys 127 (1990) 425–432.
21. MacKay RS, Supplement to "Chaos in three physical systems, Section 3: Ergodic Pumping", http://www.maths.warwick.ac.uk/~mackay/publications.html
22. Hill TL, Free energy transduction in biology (Academic, 1977).
23. Smith DA, Geeves MA, Strain-dependent cross-bridge cycle for muscle, parts I and II, Biophys J 69 (1995) 524–537, 538–552.