Internship M1
Mathematical analysis of the arrow of time

Louis Detzen

Ecole Polytechnique
Eberhard Karls Universität Tübingen
March 28th - July 22nd

Abstract

Explaining the arrow of time, ie. the apparent contradiction between the irreversibility of the macroscopic world and the reversibility of the microscopic one has been a fundamental problem since the development of thermodynamics in the 18th century. Since then, different explanations have been proposed, the most famous one being the answer given by Ludwig Boltzmann, who made the assumption that our universe was initially in a very specific state, a state of low entropy. This assumption has raised many debates among the scientific community and our present work is to study new models that try to get rid of this assumption.

Contents

1 Acknowledgments **3**

2 Introduction **3**

3 Explaining the 2nd law with Boltzmann's works **5**
3.1 The Maxwellian distribution is the typical empirical distribution 5
3.2 Derivation of the Boltzmann equation 8
3.3 The H-theorem and Landford's theorem explaining the 2nd law 11

4 Why do we need to make the assumption of special initial conditions ? **14**

5 Explaining an initially low entropy-state **16**
5.1 Boltzmann's fluctuation hypothesis 16
5.2 The Caroll-Chen scenario 17

6 Several attempts to derive an arrow of time without specific initial conditions **18**
6.1 A unbounded grow of entropy in astrophysics : Antonov's model with a star system 18
6.2 A stars' system on a torus : a numerical experiment 18
6.3 Using relativity to get an arrow of time ? 19
6.4 Another astrophysical system : a disc of accretion rotating around a binary system of stars 23

7 Conclusion **31**

A Python code for the numerical experiment **32**

B Python code to plot the tidal waves **36**

1 Acknowledgments

I would like to thank my tutor, Prof. Dr. Peter PICKL, who supervised me during the whole internship with a great kindness, as well as Prof. Dr. Roderich TUMULKA, who helped me to understand the fundamental ideas and tools of statistical mechanics.

2 Introduction

Statistical physics deals with the description of systems consisting of a large amount of objects (such as atoms, molecules, etc.) - typically more than 10^{20} [1]. One doesn't want to look at the individual behaviour of every component, because due to the huge size of the system, it would be impossible in practice, but also because in order to get a fair idea of particular features of the system, one doesn't actually need to have this full knowledge.
One distinguishes the macroscopic scale from the microscopic one. At the microscopic scale, the system is fully described by a point in a phase space, which contains all the positions and impulsions of each particle. On the macroscopic scale, one looks at macroscopic variables, such as the temperature T, the pressure P, etc., which are functions on phase space and tend to have (approximately) the same value for a lot of microscopic configurations. One calls such functions *coarse-grained* functions.

The problem that appears at this stage of description is that there is a paradox between the macroscopic laws - which are for example the laws of thermodynamics - and the microscopic ones, which are Newton's laws (or the quantum mechanical laws, but in this report, I'll only focus on the classical case). On the one hand, microscopic laws are reversible. If $x(t) = (q(t), p(t)) \in \mathbf{R}^{6N}$ is a point in the phase space Γ describing the microscopic state of the system at time t (with N the number of components and $q(t), p(t)$ containing respectively all the positions and impulsions), then Hamilton's equations read :

$$\begin{cases} \dot{q} &= \dfrac{\partial H}{\partial p} \\ \dot{p} &= -\dfrac{\partial H}{\partial q} \end{cases} \tag{1}$$

where H is the Hamiltonian of the system. Let $T(t, x)$ be the associated flow, that is, one has, $x(t) = T(t - t', , x(t'))$, or $x(t) = T^{t-t'}\, x(t')$. Then, the dynamic is reversible, which means that $T^t I T^t x = Ix$, where $I(q, p) = (q, -p)$ reverses the impulsions. Thus, both trajectories, the original trajectory and its reversed counterpart, exist.
On the other hand, macroscopic laws are irreversible. Indeed, the 2nd law states that for a closed system, the thermodynamical entropy is non-decreasing. Then, one observes only a single trajectory among the latter two, which is the "physical" one, which leads to such a growth. This time-asymmetry that appears at the macroscopic scale, although the microscopic equations of motion are reversible, is called the "arrow of time".

Boltzmann's major contribution to solve this apparent contradiction was to show that the explanation may lie in the *initial conditions* and not in the equations of motion themselves. Let's introduce some other notations to explain his reasoning. Considering a macro-variable F [2] (which one supposes determining the macro-state), at time t, one observes F_t (which value defines a macro-state ν_t). Let $\overline{\Omega}_t$ be the set of all microscopic configurations compatible with ν_t and $\Omega_t \subset \overline{\Omega}_t$ the subset that leads F_t to behave according to the macroscopic laws (e.g. the diffusion equation if F is the temperature).

The first key assumption of Boltzmann is to state that Ω_t is such that $|\Omega_t| \approx |\overline{\Omega}_t|$. This defines the set of *typical solutions*. With this assumption, one can derive the growth of Boltzmann entropy for a closed system, and thus the 2nd law of thermodynamics, if we assume these entropies to be equal. Indeed, for a given $t_1 > 0$, one has $T^{t-t_1}(\Omega_{t_1}) \subset \Omega_t$ for all $t > t_1$ by definition of Ω_t. But, one has :

$$|T^{t-t_1}(\Omega_{t_1})| = |IT^{t-t_1}(\Omega_{t_1})| << |\Omega_t|$$

as $IT^{t-t_1}(\Omega_{t_1}) \notin \Omega_t$ because it corresponds to the reversed "unphysical" behavior. One also has $|T^{t-t_1}(\Omega_{t_1})| = |\Omega_{t_1}|$ by Liouville's theorem, so one ends up with :

$$|\Omega_{t_1}| << |\Omega_t| \qquad \forall t > t_1$$

Thus, there is an exponential growth of the sets' measure until one reaches an *equilibrium*, which is by definition the set Ω_{eq} which occupies the largest part of phase space : $|\Omega_{eq}| \approx |\Gamma|$.

The Boltzmann entropy, which reads $S(t) = k \log |\Omega_t|$ (where k is Boltzmann's constant), is therefore increasing up to $S_{eq} := k \log |\Omega_{eq}|$, and such an evolution explains the "arrow of time".

But the problem doesn't end here. The latter explanation raises the following question : if the equilibrium state is the typical state in the space space, why is our universe not in a state of equilibrium ? To solve this new problem, Boltzmann makes a second assumption : the universe at the origin was in a particular state, a state of low-entropy. This second assumption then explains why, since then, the entropy of the universe has increased and will still increase in the future, until it reaches a state of equilibrium.

The present work tries to study new models that try to get rid of this assumption, which doesn't look satisfactory, because it leaves the following questions unanswered : why and how could the universe at its initial stage be in such an atypical state ? The models try to abandon this hypothesis, and we've found at the end a model that reveals an arrow of time, without specifying special initial conditions

3 Explaining the 2nd law with Boltzmann's works

In this section, we'll go over classical results derived by Boltzmann to explain the 2nd law of thermodynamics, at least for the case of the ideal gas or a rarefied gas.
First, we'll show the Maxwellian distribution (for velocities) corresponds to the dominant macro set (in a sense that we'll specify below) and maximizises the entropy.
Then, we'll derive, at least heuristically, the Boltzmann equation.
Finally, we'll show the H-theorem, which gives the growth of the entropy, for typical micro states, thanks to Lanford's theorem, and thus gives a mathematical proof of the 2nd law of thermodynamics (at least for the ideal or rarefied gas and for short times).

3.1 The Maxwellian distribution is the typical empirical distribution

To give the proof of this statement, we'll first have to show the Gaussian distribution is asymptotically a marginal of the uniform distribution. More precisely, we have :

Lemma 1 *Let $d > k \geqslant 1, \sigma > 0$, $\rho_G : \mathbf{R}^k \to [0, \infty)$ the density of $\mathcal{N}^k(0, \sigma^2 I)$.*
Let $(X_1, \dots, X_d) \sim u_R^{d-1}$ with $R = \sigma\sqrt{d}$.
Then, the marginal of $X_1, \dots X_d$ is absolutely continuous in $\mathbf{R}^k$ with density $\rho_{k,d}$ and :

$$\|\rho_{k,d} - \rho_G\|_{L^1(\mathbf{R}^k)} \to_{d\to\infty} 0$$

Proof.
1st step : we'll show that if ρ_d and ρ are normalized densities in $\mathbf{R}^d$, if $\|\rho_d - \rho\|_{L^\infty} \to_{d\to\infty} 0$, then $\|\rho_d - \rho\|_{L^1} \to_{d\to\infty} 0$
Let $\epsilon \in (0, 1)$ and B a ball centered at 0 such that $\int_B \rho \geqslant 1 - \epsilon$. By uniform convergence, one has :

$$\|\rho_d - \rho\|_{L^\infty} \leqslant \frac{\epsilon}{|B|}$$

Thus,

$$\int_B |\rho_d - \rho| \leqslant \epsilon$$

This gives $\int_B \rho_d \geqslant \int_B \rho - \epsilon \leqslant 1 - 2\epsilon$ and $\int_{B^c} \rho_d \leqslant 2\epsilon$ We get :

$$\begin{aligned}\int_{B^c} |\rho - \rho_d| &\leqslant \int_{B^c} (\rho + \rho_d) \\ &\leqslant \epsilon + 2\epsilon \\ &\leqslant 3\epsilon\end{aligned}$$

Finally, we have :

$$\int_{R^d} |\rho_d - \rho| \leqslant 4\epsilon$$

2nd step : we'll show the marginal of u_R^{d-1} converges uniformly to the Gaussian distribution.
We first admit that this marginal is absolutely continuous and has the density $\rho_{k,d}$ proportional to (ie, up to a normalization constant) :

$$\rho'_{k,d}(x) = \mathbf{1}_{x^2 \leqslant \sigma^2 d} \left(1 - \frac{x^2}{\sigma^2 d}\right)^{-k/2-1} \left(1 - \frac{x^2/2\sigma^2}{d/2}\right)^{d/2}$$

Actually, we first show $\rho'_{k,d}$ converges uniformly to $\rho'(x) = e^{-x^2/2\sigma^2}$ in R^k.
Let recall $\left(1 + \frac{x}{n}\right)^n \to_{n\to\infty} e^x$ uniformly on every compact subset of $\mathbf{R}$.
Let $\epsilon \in (0,1)$ and $r > 0$ such that $e^{-r^2/2\sigma^2} < \frac{\epsilon}{3}$. For $|x| < r$, there is uniform convergence, according the last fact.
For $|x| \geqslant r$, $\rho'(x) \leqslant \frac{\epsilon}{3}$ and :

$$\begin{aligned}
\rho'_{k,d}(x) &\leqslant \rho_{k,d}(|x|, 0, \ldots, 0) \\
&\leqslant \rho_{k,d}(r, 0, \ldots, 0) \\
\text{[pointwise convergence]} \qquad &\leqslant \rho'(r, 0 \ldots, 0) + \frac{\epsilon}{3} \\
&\leqslant \frac{2\epsilon}{3} \\
\rho'(x) &\leqslant \rho'(|x|, 0, \ldots, 0) \\
&\leqslant \frac{\epsilon}{3}
\end{aligned}$$

Thus, we get,

$$\begin{aligned}
|\rho_{k,d}(x) - \rho(x)| &\leqslant \rho_{d,k}(x) + \rho(x) \\
&\leqslant \epsilon
\end{aligned}$$

Bringing together the two cases, we get there is uniform convergence on $\mathbf{R}^k$.
Using the first step, we have that there is also convergence of the normalization constant of $\rho'_{k,d}$ to the one of ρ', which finally gives $\rho_{k,d}$ converges uniformly to ρ_G. □

Proposition 1 *Let $X = (X_1, \ldots, X_d)$ with uniform distribution u_R over $\mathcal{S}_R^{d-1}$, the sphere in $\mathbf{R}^d$ of radius $R = \sigma\sqrt{d}$. Let $u^{emp} := \frac{1}{d}\sum_{k=1}^d \delta_{X_k}$, $u_G \sim \mathcal{N}^d(0, \sigma^2 I)$, and a partition $(A_1 \ldots, A_r)$ of $\mathbf{R}$ ($r \geqslant 1$).*
Then, for all $\epsilon > 0$, we have :

$$\mathbb{P}(\forall i : |u^{emp}(A_i) - u_G(A_i)| < \epsilon) \to_{d\to\infty} 1$$

Proof.
Let $F_i := u^{emp}(A_i) = \frac{1}{d}\sum_{k=1}^d \mathbf{1}_{X_k \in A_i}$ and $P_i := \mathbb{P}(X_1 \in A_i)$. Then $\mathbb{E}F_i = P_i$.

Let show $VarF_i \to_{d\to\infty} 0$

$$\begin{aligned}
VarF_i &= Cov(F_i, F_i) \\
&= \frac{1}{d^2}\sum_{j,k=1}^{d} Cov(\mathbf{1}_{X_j\in A_i}, \mathbf{1}_{X_k\in A_i}) \\
&= \frac{1}{d^2}\sum_{j=1}^{d} Var(\mathbf{1}_{X_j\in A_i}) + \frac{1}{d^2}\sum_{j\neq k} Cov(\mathbf{1}_{X_j\in A_i}, \mathbf{1}_{X_k\in A_i}) \\
&\leqslant \frac{P_i(1-P_i)}{d} + \frac{d(d-1)}{d^2}|Cov(\mathbf{1}_{X_1\in A_i}, \mathbf{1}_{X_2\in A_i})| \\
VarF_i &\leqslant \frac{P_i(1-P_i)}{d} + |\mathbb{P}(X_1\in A_i, X_2\in A_i) - P_i^2|
\end{aligned}$$

But, $P_i \to_{d\to\infty} \mu_G(A_i)$ and $\mathbb{P}(X_1 \in A_i, X_2 \in A_i) \to \mu_G(A_i)^2$ according to the previous lemma applied for $k = 1, 2$. Thus, we get $VarF_i \to_{d\to\infty} 0$.

Now, let $\epsilon > 0$ and d large enough such that $|P_i - \mu_G(A_i)| < \frac{\epsilon}{2}$. Then, by Chebyshev inequality, we have :

$$\begin{aligned}
\mathbb{P}\left(\bigcup_i \{|F_i - \mu_G(A_i)| \geqslant \epsilon\}\right) &\leqslant \mathbb{P}\left(\bigcup_i \{|F_i - P_i| + |P_i - \mu_G(A_i)| \geqslant \epsilon\}\right) \\
&\leqslant \sum_{i=1}^{r} \mathbb{P}\left(|F_i - P_i| \geqslant \frac{\epsilon}{2}\right) \\
&\leqslant \sum_{i=1}^{r} \frac{4VarF_i}{\epsilon^2} \to_{d\to\infty} 0 \qquad \square
\end{aligned}$$

Let's apply this result to the case of the ideal gas, composed of N identical particles of mass m, confined in a compact subset $\Lambda \subset \mathbf{R}^3$. The (ordered) phase space is $\Gamma = (\Lambda \times \mathbf{R}^3)^N$ and the Hamiltonian H is $H = \sum_{i=1}^{N} \frac{p_i^2}{2m}$. As the total energy is conserved, the phase space to consider is rather the energy shell $\Gamma_E = \mathcal{S}^{3N-1}_{\sqrt{2mE}}$, where $E = \dfrac{e}{n}$ is the total energy and e is the average energy per particle. We omitted the Cartesian product, as the positions play no role, except in the computation of the entropy. Then, the last statement says that for most phase points on the sphere (relative to the Liouville measure on the energy shell), $u^{emp} \sim \mathcal{N}^d(0, \sigma^2 I)$ when $N >> 1$ (which is the case as $N \approx 10^{23}$ particles usually). Here, $\sigma^2 = \frac{2mE}{3N} = \frac{2}{3}me$. As $e = \frac{3}{2}kT$, by the equipartition theorem, where T is the temperature of the gas, one gets $\sigma^2 = mkT$ and :

$$u^{emp}(x) \approx \frac{1}{N!|\Lambda|^N\sqrt{2\pi mkT}} e^{-p^2/mkT}$$

where the prefactor $\frac{1}{N!|\Lambda|^N}$ takes into account the particles are identical and that the true phase space contains positions.

This result also applies for **the hard sphere gas model** where particles are like billiard balls which interact through elastic collisions (ie, the energy and the impulsion of two particles is conserved for each collision). We also suppose that the **gas is "rarefied"**, ie has low density (in a sense that precise later). With this model, Boltzmann used the hypothesis of molecular chaos ("Stosszahlansatz"), which says velocities are uncorrelated (and independent of position). The phase points are then distributed according to the uniform distribution on the sphere and the last result still holds, as for a given time, between collisions, the Hamiltonian energy has the same expression.

Let's do the link with the macro states.
Let $f = (f_i)_{i=1}^r$ be a well-chosen coarse-graining of u^{emp}, associated to a partition of $\mathbf{R}^3$. This macro-variable (which Boltzmann used) defines a partition of the total phase space into macro sets Γ_f. Then, we define

$$(f_i)^{eq} := \frac{1}{N!|\Lambda|^N\sqrt{2\pi mkT}} \int_{A_i} e^{-p^2/mkT}\, dp$$

and in this equality, one has to round the value, with respect to a resolution Δf (to have a finite partition of the phase space)

The last result gives that

$$|\Gamma_{eq}| \approx |\Gamma_E|$$

With this choice of macro-variables, one can also be more precise and show the different in size is like :

$$\frac{|\Gamma_{eq}|}{|\Gamma_E|} \approx 1 - e^{-\alpha N}$$

where $\alpha > 0$ is some constant (related to the entropy per particle of the equilibrium state).
Thus, the macro state associated to the Maxwellian distribution is a dominant macro state and represents the thermodynamic equilibrium. As $S(f) = k\log(|\Gamma_f|/N!)$, it maximizes the entropy.

3.2 Derivation of the Boltzmann equation

Above we explored the structure of phase space for the hard sphere gas model, and showed that a dominant macro state, ie equilibrium, exists and maximizes the entropy. We now would like to explore the dynamic of a phase point, in particular to know the evolution of its entropy with time. This is the role of the Boltzmann equation which is a continuous approximation of the evolution of the macro state for typical phase points of an initial macro set (we'll see in which sense typicality plays a role here with Landford's theorem below).
We consider the Boltzmann-Grad limit, which means that the gas is "rarefied" and more precisely :

$$N \to \infty,\ a \to 0,\ 4Na^2 \to \lambda$$

where a is the radius of the spheres and λ is some positive constant.
Indeed, the volume occupied by the spheres $N\frac{4\pi}{3}a^3$ tends to 0 in this limit.
The Boltzmann equation reads :

$$\left(\frac{\partial}{\partial t} + v \cdot \nabla_q\right) f(q,v,t) = Q(q,v,t)$$

where the collision term Q is given by :

$$Q(q,v,t) = \lambda \int_{\mathbf{R}^3} dv_\star \int_{\mathcal{S}^2} d\omega \, \mathbf{1}_{\omega\cdot(v-v_\star)>0} \, \omega\cdot(v-v_\star)(f(q,v',t)f(q,v'_\star,t) - f(q,v,t)f(q,v_\star,t)) \quad (2)$$

where $\mathcal{S}^2$ is the 2d unit sphere.
The speeds $v', v'_\star$ are the speeds for two particles (1 and 2 respectively) after one elastic collision and are given by :

$$v' = v - [(v-v_\star)\cdot\omega]\omega$$
$$v'_\star = v_\star + [(v-v_\star)\cdot\omega]\omega$$

and ω is related to the positions of the two particles by :

$$\omega = \frac{q_2 - q_1}{2a}$$

This follows from the conservation of energy and momentum as the collision is supposed to be elastic.
In addition to the differential equation, there is a boundary condition of "specular reflection", ie for $(q,v) \in \partial\Lambda \times \mathbf{R}^3$,

$$f(q,v,t) = f(q, v - 2(v\cdot n)n, t)$$

where n is the outward unit normal vector at q.

The term in the left of the equation is the classic term of an transport equation, in the absence of collisions, so the task is to derive (at least heuristically) the expression for the collision term. The later is given by :

$\boxed{NQ(q,v)dq\,dv = \text{rate of change in particle number in } dqdv = \text{gain - loss}}$

In the computation we write infinitesimal volumes, but there are supposed to contain many particles, so in practice we only take the continuous limit at the end of the derivation.

First, let's compute the loss in particles. Let $\delta\mathcal{N}$ be the number of collisions within dt of 2 balls in dq with velocities in dv and $dv_\star$ and collision parameter in $d\omega$. For each first ball in $dqdv$ whose number is $Nf(q,v)dqdv$, there are as much collisions (and thus losses) as 2nd balls that can collide in dt. This number is equal to $Nf(q,v_\star)d_{v\star}dC$, where dC is the volume of the collision cylinder $d\mathcal{C}$, when the 1st ball is fixed. In the frame of reference of the 1st ball (ie in which $v=0$), we get $dC = -4a^2 v_\star\cdot\omega d\omega dt$ (see figure below).

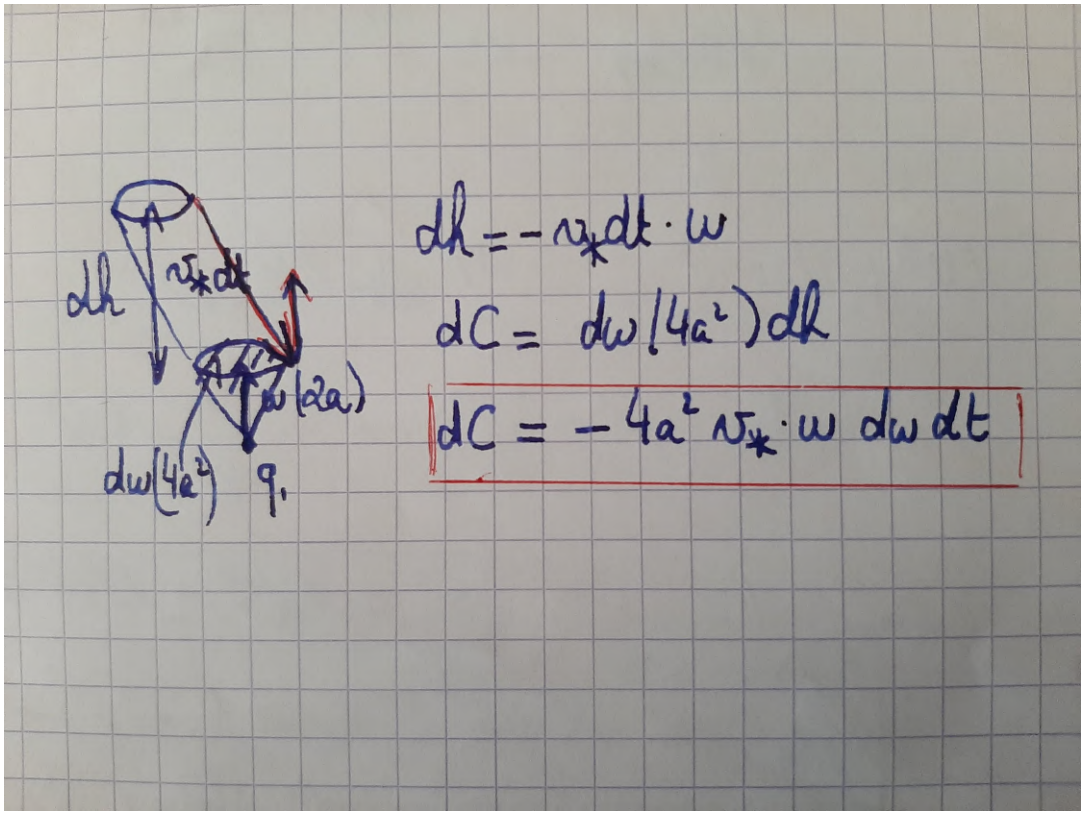

Figure 1: Collision cylinder (with respect to the 1st ball

By Galilean transformation, in the laboratory frame, we have $dC = 4a^2(v - v_\star) \cdot \omega d\omega dt$. Multiplying these numbers by $\mathbf{1}_{(v-v_\star)\cdot\omega>0}$, which expresses the condition that is needed to have a collision, we get :

$$\begin{aligned}\text{loss} &= \int_{\omega\in\mathcal{S}^2}\int_{v_\star\in\mathbf{R}^3} \delta\mathcal{N} \\ &= 4a^2N^2\int_{\mathcal{S}^2}\int_{\mathbf{R}^3} dtdqdvdv_\star d\omega\, \mathbf{1}_{\omega\cdot(v-v_\star)>0}\,\omega\cdot(v-v_\star)f(q,v,t)f(q,v_\star,t)\end{aligned}$$

In the computation, we multiplied the number of particles for the 1st ball by the number of particles for the 2nd ball (taking into account that a collision takes place), treating them separately. We used the hypothesis of molecular chaos, which we've already seen before.

For the computation of the gain, let $R_w : (v, v_\star) \to (v', v'_\star)$ the operator which gives the speeds post-collisions provided the initial speeds of the two particles at the time just before the collision. We'll use the following properties of this operator :

1. $R_\omega \in O(6)$

2. $\det R_\omega = -1$

3. $R_\omega^2 = I$

4. $R_{-\omega} = R_\omega$

5. $\omega \cdot (v' - v'_\star) = -\omega \cdot (v - v_\star)$

Let $C \subset \mathbf{R}^6$. The gain of particles due to collisions of the considered particles and such that $u' = (v', v'_\star) \in C$ is given by :

$$\begin{aligned}
\text{gain}_C &= \int_{\omega \in \mathcal{S}^2} \int_{u \in \mathbf{R}^6} \delta\mathcal{N} \\
&= 4a^2N^2 \int_{\mathcal{S}^2} \int_{\mathbf{R}^6} dtdqdud\omega \, \mathbf{1}_{\omega\cdot(v-v_\star)>0} \, \omega \cdot (v - v_\star) f(q, v, t) f(q, v_\star, t) \, \mathbf{1}_{R_\omega u \in C} \\
[u' = R_\omega u, u' \Leftrightarrow u] &= 4a^2N^2 \int_{\mathcal{S}^2} \int_{\mathbf{R}^6} dtdqdud\omega \, \mathbf{1}_{\omega\cdot(v'-v'_\star)>0} \, \omega \cdot (v - v_\star) f(q, v', t) f(q, v'_\star, t) \, \mathbf{1}_{u \in C} \\
&= 4a^2N^2 \int_{\mathcal{S}^2} \int_{C} dtdqdud\omega \, \mathbf{1}_{\omega\cdot(v-v_\star)<0} (-\omega) \cdot (v - v_\star) f(q, v', t) f(q, v'_\star, t) \\
[\omega \mapsto -\omega] &= 4a^2N^2 \int_{\mathcal{S}^2} \int_{C} dtdqdud\omega \, \mathbf{1}_{\omega\cdot(v-v_\star)>0} \, \omega \cdot (v - v_\star) f(q, v', t) f(q, v'_\star, t)
\end{aligned}$$

The researched gain is gain $=$ gain$_{dv \times \mathbf{R}^3}$, thus we have :

$$\text{gain} = 4a^2N^2 dqdvdt \int_{\mathcal{S}^2} \int_{R^3} dv_\star d\omega \, \mathbf{1}_{\omega\cdot(v-v_\star)>0} \, \omega \cdot (v - v_\star) f(q, v', t) f(q, v'_\star, t)$$

and this gives the expression of $Q(q, v, t)$ according to (2).

3.3 The H-theorem and Landford's theorem explaining the 2nd law

In his work, Boltzmann defined the H functional by :

$$H(f) = \int_\Lambda \int_{\mathbf{R}^3} dqdv \, f \log f$$

The H-theorem states that $H(f(t, \cdot, \cdot))$ is non-decreasing where $f(t, \cdot, \cdot)$ is a solution of the Boltzmann equation. The link with the 2nd law is that for the hard sphere rarefied gas model, one has :

$$s(f) = -kH(f)$$

where $s(f)$ is the entropy per particle corresponding to the macro state f. This equality can be derived by taking the limit $N \to \infty$ and then the continuous approximation, ie. the size of the cells in phase space $dqdv$ goes to 0.
In other words, let show the the following theorem :

Theorem 1 *If $f_t = f(t,\cdot,\cdot)$ is a solution of the Boltzmann equation, then :*

$$\frac{dH(f_t)}{dt} \leqslant 0$$

Proof.
Let's introduce some notations : $f' := f(q,v'), f_\star := f(q,v_\star), f'_\star := f(q,v'_\star)$
First, let state the lemma :

Lemma 2 *For some test function $\phi(q,v)$, one has :*

$$\int_{\mathbf{R}^3} dvQ\phi = \frac{\lambda}{4}\int_{\mathbf{R}^6} dvdv_\star \int_{\mathcal{S}^2} \mathbf{1}_{\omega\cdot(v-v_\star)>0}\,\omega\cdot(v-v_\star)(f'f'_\star - ff_\star)(\phi+\phi_\star-\phi'-\phi'_\star)$$

This follows directly from the symmetries of the operator R_ω which leaves the integral unchanged. For example, the change of variables $v \Leftrightarrow v_\star$ and then $\omega \mapsto -\omega$ gives the term in $\phi_\star$, likewise for the other terms. □

Secondly, let show the "Boltzmann inequality" :

Proposition 2 *If $f(q,v) > 0$, then*

$$\int dv\, Q\log f \leqslant 0$$

We use the lemma above with $\phi = \log f$, which gives :

$$\int dv\, Q\log f = \frac{\lambda}{4}\int_{\mathbf{R}^6} dvdv_\star \int_{\mathcal{S}^2} \mathbf{1}_{\omega\cdot(v-v_\star)>0}\,\omega\cdot(v-v_\star)\underbrace{(f'f'_\star - ff_\star)\log\left(\frac{ff_\star}{f'f'_\star}\right)}_{\leqslant 0} \leqslant 0$$

as for any $y,z>0$, one has : $(z-y)\log\frac{y}{z} \leqslant 0$
Let's now show the H-theorem, by computing $\frac{dH}{dt}$:

$$\begin{aligned}
\frac{dH}{dt} &= \frac{d}{dt}\int dq \int dv f\log f \\
&= \int dqdv\,\partial_t(f\log f) \\
&\int dqdv\,(\partial_t f)(1+\log f) \\
&= \int dqdv\,(-(1+\log f)v\cdot\nabla_q f + Q(1+\log f)) \\
&= -\int dqdv\,(1+\log f)v\cdot\nabla_q f + \int dqdv\,Q + \int dqdv\,Q\log f
\end{aligned}$$

But,

$$\int dqdv\, Q = 0$$

which comes from the lemma (2) applied with $\phi = 1$

$$\int dqdv\, Q \log f \leqslant 0$$

according to the "Boltzmann inequality" and,

$$\begin{aligned}
\int dqdv\, (1 + \log f) v \cdot \nabla_q f &= \int dqdv\, v \cdot \nabla_q (f \log f) \\
&= \int_\Lambda dq\, \nabla_q \left(\int_{\mathbf{R}}^3 dv\, v f \log f \right) \\
\text{[Gauss theorem]} \qquad &= \int_{\partial\Lambda} d\sigma(q) \int_{\mathbf{R}}^3 dv\, n \cdot v f \log f \\
&= 0
\end{aligned}$$

as $n \cdot v$ is antisymmetric in the n-coordinate and the "specular reflection" implies that f is symmetric in the n-coordinate.

As a result, one finally gets : $\boxed{\dfrac{dH}{dt} \leqslant 0}$ $\square$

At this stage, the H-theorem is not enough to derive the 2nd law, because we don't know *a priori* f_t corresponds the empirical distribution of $x(t)$ (except if we trust the heuristic derivation made before). This is the role of Lanford's theorem which reads :

Theorem 2 *For large N, for almost $x(0) \in \gamma_{f_0}$ (ie with empirical distribution f_0), the emprical distribution of $x(t)$ is close to f_t, where f_t is the solution of the Boltzmann equation, at least for $t \in [0, \frac{1}{5}\bar{t})$, where $\bar{t}$ is the mean free time between two collisions.*

Although there is no mathematical proof up to now, it is believed that Lanford's theorem applies for larger times, at least up to recurrence times.

Applying this theorem, one gets the 2nd law of thermodynamics for this particular model of gas, which states that its entropy should be non-decreasing.

4 Why do we need to make the assumption of special initial conditions ?

This question comes from a problem emerging from Lanford's theorem. It is valid for positive times, but one can look at the evolution of f_t for negative times if one uses $(f_0)^r(q,v) = f(q,-v)$, as the microscopic laws (ie Newton laws) are reversible. But then, one finds $(f_0)^r$ is a solution of the Boltzmann equation with $-\lambda < 0$, and in this case $\frac{dH}{dt} \geqslant 0$ and the entropy is non-increasing (see figure).

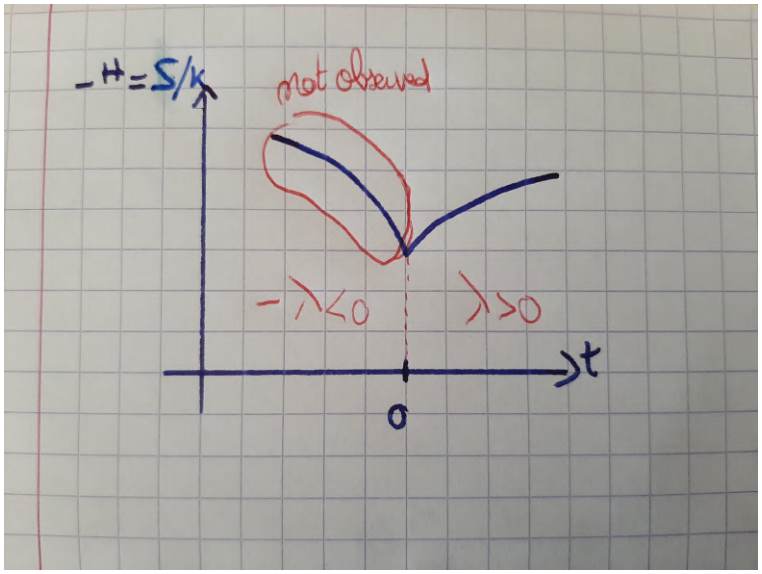

Figure 2: Entropy curve derived from Lanford's theorem

This part of the entropy curve is not observed in reality, so one has to add a postulate to the mechanical laws :

> A closed real-world system that at time t_0 has macro-state ν appears macroscopically in the future, but not the past, of t_0 like a system that at time t_0 is in a typical micro-state compatible with ν

(This corresponds to Basic fact 2 in [1])

This postulate states that there is a unique arrow arrow of time (and not two as suggested by Lanford's theorem) pointing towards the future. This is more refined than the 2nd law (which implies the existence of an arrow of time), because it is a statement in the framework of statistical mechanics, from which one can derive a statement about the irreversibility of thermodynamics.

From this, one would like to give an explanation of this fact, and one possibility is to make the assumption that our universe was set in very particular initial conditions at the time of the Big Bang. More precisely, one can make the following assumption :

> The real world as a whole (i.e., the universe) appears macroscopically like a system that at the big bang is in a typical micro-state compatible with the universe's macro-state at the big bang.

(This corresponds to Basic fact 3 in [1])

That is to say, the initial phase point of the universe is considered to be typical in its initial macro state. Thus, one can apply Lanford's theorem to it, which gives that the current phase point of the universe is no more typical in its current macro state, but it behaves like the typical points of it. This is exactly what the previous postulate says. The other part of the curve given by Lanford's theorem is not observed, as there is no previous time before the Big Bang (if one considers it as the beginning of time, which is another assumption).

5 Explaining an initially low entropy-state

The first way to deal with this problem of having to explain why the universe was in such an atypical state of low-entropy, is to consider it as a fact, in addition to the fundamental laws of physics. This approach is called the **Past Hypothesis**. But there has been other attempts to deal with this problem, without treating it as a postulate.

5.1 Boltzmann's fluctuation hypothesis

In this explication due to Boltzmann, one makes the following reasoning : he makes first the assumption that time is eternal and secondly that most of the time, it lies in a state of thermal equilibrium, corresponding to a maximum value of the universe's entropy. But, since the Big Bang, our universe hasn't been in equilibrium. Boltzmann argues, that in this model of universe being eternal, due to the Poincare's recurrence theorem, there might be at some rare times fluctuations of the entropy that bring it to a low value, the deeper the fluctuation is, the more improbable it could occur. From this state of reasoning, there are 2 possibilities to explain the history and current state of our universe :

1. Either a fluctuation F_0 occurred that caused the universe to be at the Big Bang in a very special state of low-entropy. Since then, we are on the upward slope, until the entropy reaches again its highest value, that would correspond to the "heat death" of the universe.

2. Or we're finding ourselves direct at the bottom of such a fluctuation, but that means that the past of the universe would appear as a fake memory, caused by this fluctuation, which seems a little absurd.

But if one still wants to hold the first line of reasoning, another problem comes, that looks like the second point. Imagine a fluctuation that would give birth to a single brain, with the same (fake) memory of the universe that we have, the rest of the universe being in thermal equilibrium. Although this situation seems more absurd than before, this fluctuation corresponds to a less deep fluctuation than F_0, thus it has a (exponentially) higher probability to occur.

Then, one would like to apply the "anthropic principle", which says that when considering probabilities, one has always to deal with conditional probabilities, where the condition requires one human exists, to be able to talk about physical theories or measurements. The probability a second human exists given the existence of the first human is then tremendously small.

Furthermore, there is a particular case of this principle, called the "Copernican principle", which states that "we should see what a typical observer (weighted with life span) in a typical universe sees" [1]. Using this principle, one is forced to conclude that typically speaking, we should be Boltzmann brains, which we are actually not, as such a brain will immediately

disintegrate. [1] Such an argument can also apply in the same way, if one considers the room one finds itself in, the rest of the universe lying in thermal equilibrium.

One could also claim that our universe is not typical, or we are not typical observers, but one doesn't want to give up easily the Copernican principle, because it is quite a natural hypothesis in science, so this "reductio ad absurdum" argument shows that Boltzmann's fluctuation hypothesis cannot work.

5.2 The Caroll-Chen scenario

Caroll and Chen use the model of cosmological inflation in their model : they assume like Boltzmann that time is eternal and the big bang corresponds to a certain space-time region [1]. On the energy level surfaces, the entropy of the universe is finite but not bounded : contrary to the previous explanations, there no highest value for the entropy corresponding to a thermal equilibrium. In such a scenario, every macro-state appears then as a "state of low-entropy", which solves the original question. But one has to abandon the use of a normalized measure, because the Liouville measure that is usually used to measure volume in phase space and thus to speak of "typicality" is now infinite, and the latter notion has to be redefined.

[1] Although its life span may be less than a second, it doesn't contradict the typicality argument, as Boltzmann brains should appear exponentially more frequent than F_0.

6 Several attempts to derive an arrow of time without specific initial conditions

6.1 A unbounded grow of entropy in astrophysics : Antonov's model with a star system

At the end of his article [3], Antonov takes a "special distribution" of a system of interacting stars. He makes the major assumption that the processes of star escapes are slow in comparison to the system's dynamic and thus uses that there is conservation of mass and energy. He considers a particular distribution where the stars are grouped into two separated parts which are at a positive distance from each other : the first part has mass αM (resp. βM, such that $\alpha + \beta = 1$) and is distributed uniformly into a sphere of radius ρ_1 (resp. ρ_2), with a uniform Maxwellian velocity distribution with c_1 being the maximum speed (resp. c_2). He then takes the limit as ρ_2 tends to 0, $\beta \to 0$ (and then $\alpha \to 1$). He shows that one gets $c_1 \to +\infty$ and the system's entropy, which is a function of c_1, grows unbounded as well.

Therefore, one obtains a model where there is an arrow of time through an unbounded increase of the entropy, as t goes to infinity, because of the maximum velocities of the stars which increase due to the gravitational interactions and under the assumption that they are located in a bounded region.

But the latter hypothesis should be reassessed : if the maximum velocities grow unbounded, the typical time for stars' escapes should not remain negligible with respect to the system's dynamic.

6.2 A stars' system on a torus : a numerical experiment

My tutor and his PhD student Aaron Schaal had considered a system of stars interacting though the gravitational force on a torus. They expected the potential energy (PE) to decrease unbounded, whereas the kinetic energy (KE) would grow unbounded, the total energy being conserved. They thought such a behavior would be observed because the system is confined, and due to the gravitational attraction, the stars would get always closer from one another. A growth of the kinetic energy would provide an arrow of time for this system. However, it was only a intuition, they couldn't provide a proof for this result.

On my side, I tried to look what a numerical experiment would give as a result. I took a Python code from Philip Mocz [4], which was written to simulate a N-body system in $\mathbf{R}^3$. The position space being a torus, I added periodic boundary conditions for each coordinate to simulate a torus and computed a "distance" function on the torus. Because of the boundary conditions, there are two possibilities to compute the distance on the torus, given the natural distance in $\mathbf{R}^3$, as illustrated below.

In the appendix, there is the code I used for the simulation (most of the comments are from the original author). I considered a system with $N = 10$ stars, one of them having a huger mass as the other ones, in order to simulate a kind of solar system, the greatest mass playing the role of the Sun.

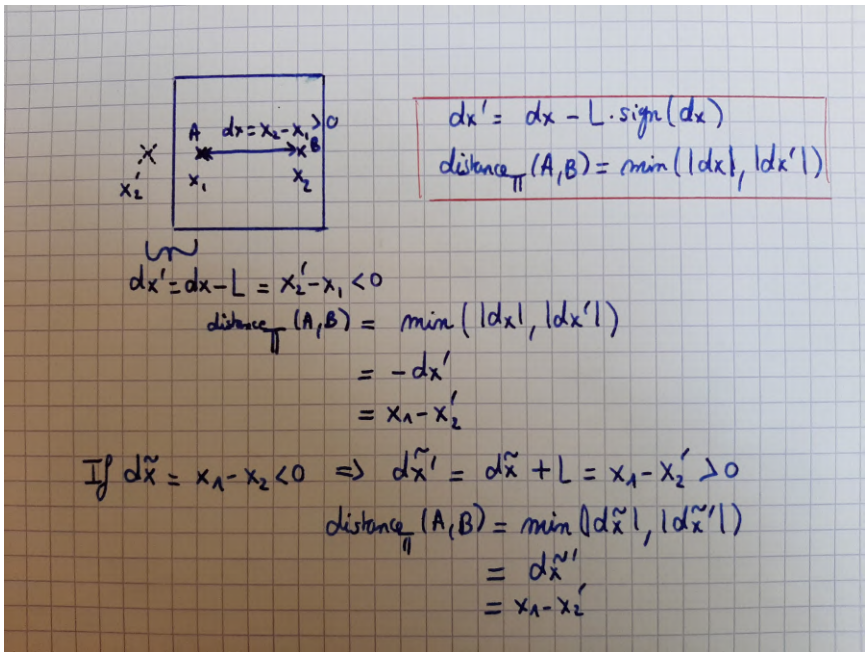

Figure 3: Formula for the distance between two stars (in one direction) on the torus

Below are the results of this numerical experiment for the parameters given in the code.

We can observe a increase of the kinetic energy with time, but we don't know if, for later times, it will tend either to a maximum (and thus the system will tend to an equilibrium) or grow unbounded as we would like to have. It is difficult to answer this question numerically, as we don't know the physical meaning of the values used for the different quantities, ie. what is the physical scale corresponding to the time, the gravitational constant, and so on. But such an experiment can give confidence in the initial intuition.

6.3 Using relativity to get an arrow of time ?

I had the idea that we might get an arrow of time if we would use the dilatation of times given by the theory of relativity.

For example, if for the observer Alice, she would observe an unbounded growth of some physical quantity in finite times, another observer, called Bob, could observe in his referential a similar unbounded growth, but in infinite times (or at least in times greater than the age of the universe).

I thought about a system of two particles in the neighborhood of a Schwarzschild black hole, falling through each other, and hitting at the event horizon, as illustrated below.

Let say Alice is attached to one of this particle. In her referential, she would measure an unbounded growth of the kinetic energy of the system, up to the time when she kicks the other particle. But the observer Bob, who is far away from the black hole, would observe that the two particles get always closer and closer, but that they will never hit each other.

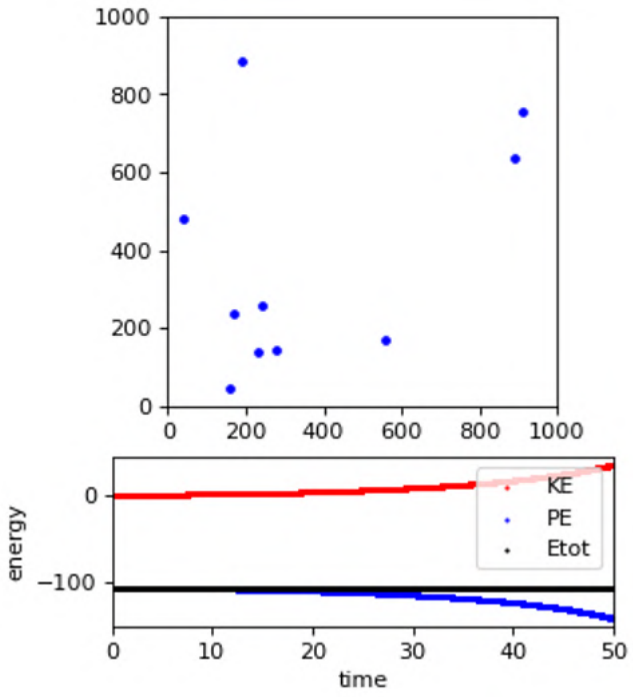

Figure 4: Simulation of a 10-star system on a torus of length $L = 1000$; E_{tot} is the total energy

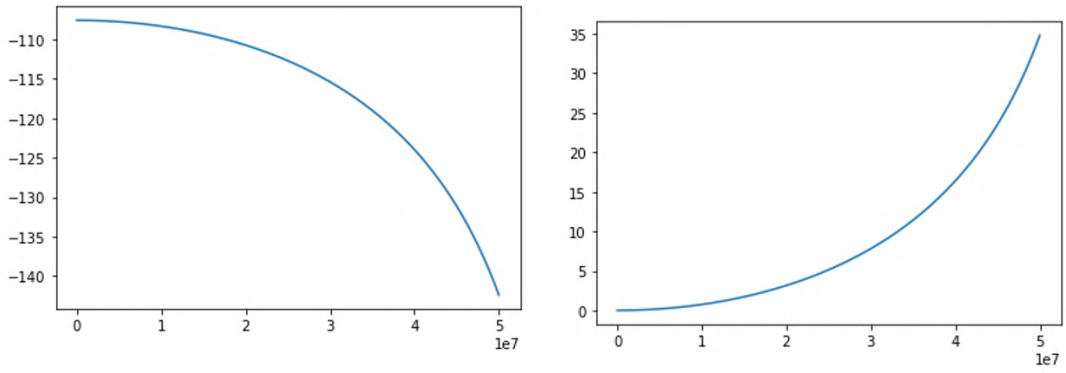

Figure 5: Evolution of the potential energy (PE) - on the left - and kinetic energy (KE) - on the right - of the system with time

Indeed, "time freezes" at the horizon for Bob, because of the time dilatation, caused by the black hole's gravity. More precisely, if $\Delta\tau$ is the time measured by Alice, the corresponding time measured by Bob Δt is given by :

$$\Delta t = \frac{\Delta\tau}{\sqrt{1 - \frac{2GM}{rc^2}}}$$

where r is the radius coordinate from the black hole's center, G the gravitational constant and M the mass of the black hole. When r tends to $r_s = \frac{2GM}{c^2}$, which is the Schwarzschild radius of the event horizon, Δt tends to infinity.

In his referential, he would measure an unbounded growth of the kinetic energy of the system, but in infinite time.

So we could conclude that we would get an arrow of time, but the problem is that the trajectory of the particles in the phase space is not typical : we have considered specific initial conditions, because we imposed the configurations (in spaces and velocities) to be symmetric.

Then, I made the following reflection : "why couldn't we consider a system located at the horizon where we would have a unbounded growth in finite times ?" In such a case, one would observe an unbounded growth in infinite times far away from the black hole, for the same reason as earlier. But, we can't consider a system a rest at the event horizon, because it has to fall into the black hole, and no finite acceleration can keep it at rest. Thus, I thought we could rather consider a similar system, outside of the black hole, at rest at a very close distance from the horizon. Here, it would work, if the times considered from the point of view of Bob are greater than the age of the universe.

But, there is still a remaining problem : one has to find a system where a physical quantity

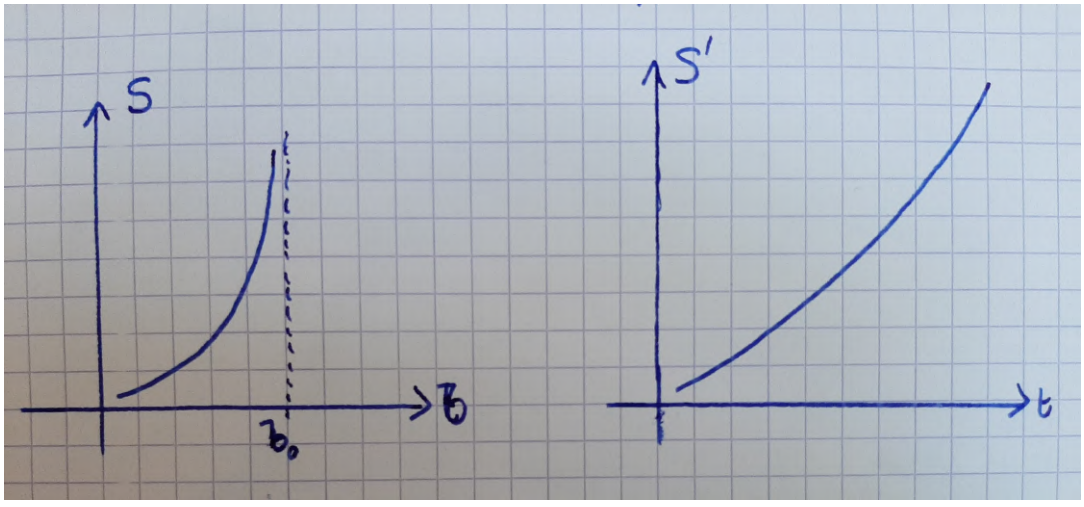

Figure 6: How to get an arrow of time with the dilatation of times

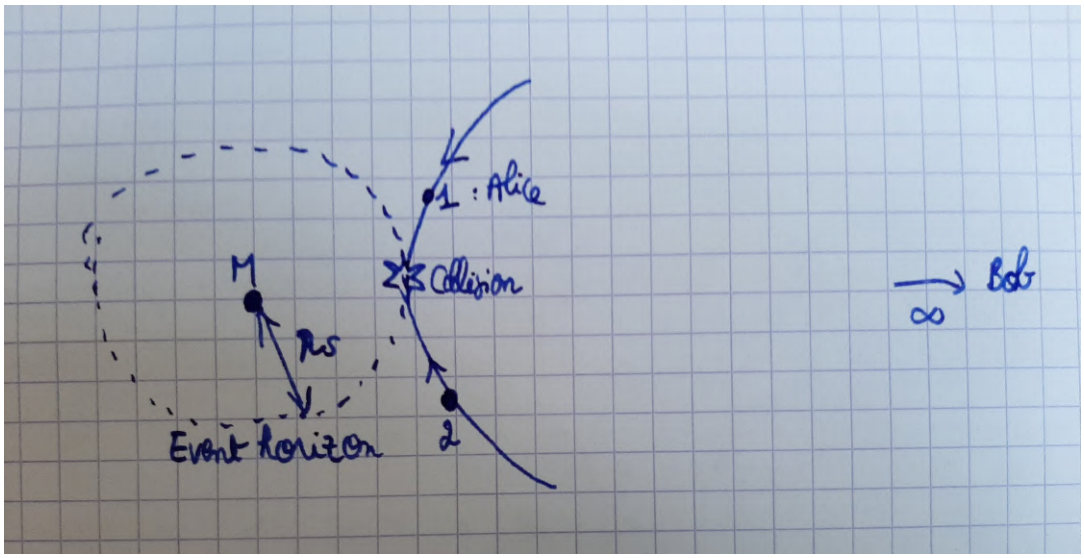

Figure 7: 2-particles system colliding at the event horizon of a black hole

(which is thermodinamically relevant) grows unbounded in finite times. I couldn't get further in this reflection. [2]

[2] I also didn't consider the relativistic transformation of the quantity which plays the role of the entropy (ie the kinetic energy here). One would also have to consider this to see what happens.

6.4 Another astrophysical system : a disc of accretion rotating around a binary system of stars

My tutor and his PhD student thought about another system, composed of a ring of a gas (like a kind of disk of accretion) rotating around a binary system of stars (which are rotating around their center of mass). The gravitational field created by this pair of stars, due to their rotation, evolves with time, which creates tidal forces, and would create a rotating wave of gas in the ring. The kinetic energy of this wave could be stored by an idealized machine, like a tidal power station on Earth which stores the energy of the waves caused by the Moon's tidal forces. In this system, the stars would get always closer from each other and would hit in infinite times. Correspondingly, the tidal wave would get always faster (but its height should get smaller) and the kinetic energy might grow unbounded as well. This is also only an intuition and I tried to make some computations to see what would happen.

The gas of particles of the ring is considered to be an ideal gas, meaning that there are no interactions between the particles, which only react to the gravitational field of the two stars. We denote by ω_0 the angular velocity of the stars rotating around each other at a distance r_0 from their center of mass (which corresponds to the origin). The stars are supposed to have the same mass M, which simplifies the computations.

Let consider infinitesimal slices of the ring of mass dm considered as a punctual particle. The ring of gas can be viewed as a rigid chain of many particles, whose radial coordinate differs from $\Delta r (\in \mathbf{R})$ from their position at rest $r = R$.

In the referential frame rotating with the stars at w_0, we consider the gravitational potential energy per unit of mass u, which is :

$$u = -GM\left(\frac{1}{s_+} + \frac{1}{s}\right)$$

where

$$\begin{aligned} s_\pm^2 &= r_0^2 + (R + \Delta r)^2 \pm 2r_0(R + \Delta r)\cos\theta \\ &= s_0^2 + 2(R \pm r_0\cos\theta)\Delta r \pm 2r_0 R\cos\theta \end{aligned}$$

where $s_0^2 = r_0^2 + R^2$ and θ is the angle of the particle in the rotating plane, with respect to the axis which joins the two stars (in the rotating frame) - see figure.

Let first compute $\dfrac{\partial s\pm}{\partial r}$: differentiating $s_\pm^2 = r_0^2 + r^2 \pm 2r_0 r\cos\theta$ with respect to r gives, $2s_\pm s_\pm = 2rdr \pm 2r_0\cos\theta$, thus :

$$\begin{aligned} \frac{partials_\pm}{\partial r} &= \frac{2r \pm 2r_0\cos\theta}{2s\pm} \\ &= \frac{r \pm 2r_0\cos\theta}{s\pm} \end{aligned}$$

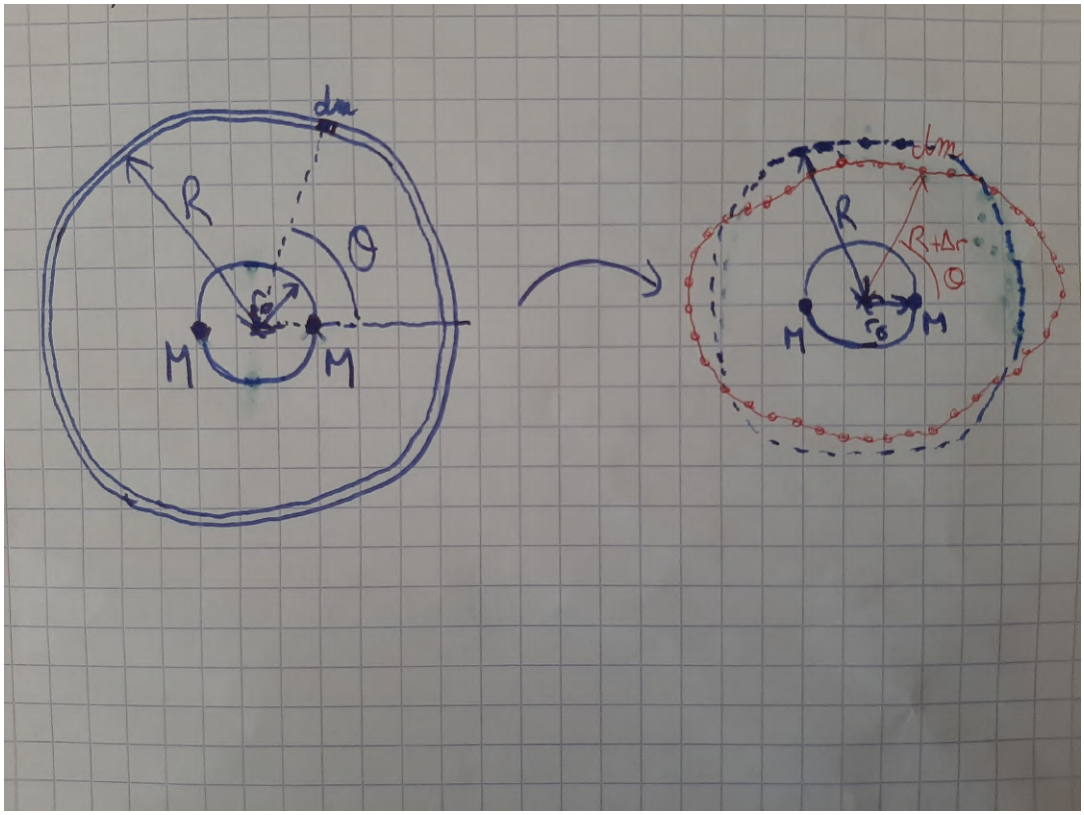

Figure 8: Binary star system and the ring of gas in the rotating frame

Thus, the radial force (per unit of mass) f_r exerted on a particle will be :

$$\begin{aligned} f_r &= -\partial u \partial r \\ &= -GM\left(\frac{R + r_0\cos\theta + \Delta r}{s_+^3(\Delta r)} + \frac{r - r_0\cos\theta + \Delta r}{s_-^3(\Delta r)}\right) \end{aligned}$$

Let suppose $\Delta r << R - r_0$, then we can do an Taylor expansion in the first order of Δr :

$$\begin{aligned} s_\pm^{-3} &= (s_0^2 \pm 2r_0R\cos\theta)^{-3/2}\left(1 - 3\frac{(R \pm r_0\cos\theta)\Delta r}{s_0^2 \pm 2r_0R\cos\theta}\right) \\ &= (s_0^2 \pm 2r_0R\cos\theta)^{-3/2} - 3(s_0^2 \pm 2r_0R\cos\theta)^{-5/2}(R \pm r_0\cos\theta)\Delta r \end{aligned}$$

We choose the unit of time such that $GM = 1$. The expression of the force will then be :

$$\begin{aligned} f_r(\Delta r, r_0) \approx &- \sum_{\sigma=\pm1}(R + \sigma r_0\cos\theta)(s_0^2 + 2\sigma r_0R\cos\theta)^{-3/2} \\ &- \Delta r \sum_{\sigma=\pm1}\left(-3(s_0^2 + 2\sigma r_0R\cos\theta)^{-5/2}(R + \sigma r_0\cos\theta)^2 + (s_0^2 + 2\sigma r_0R\cos\theta)^{-3/2}\right) \end{aligned}$$

The radial acceleration of the particle a_r is :

$$a_r(\Delta r, r_0) = -(\omega - \omega_0)^2(R + \Delta r)$$

where ω_2 is the angular velocity of the particle in the inertial frame / non-rotating frame (which should be approximately equal for all particles, as the particles are linked together by the rigid chain, which generates a tangential force between neighbours). The inertial forces (the centrifugal force and the Coriolis force) are taken into account in this formula and are considered as acceleration terms.

There is a constraint on the value of ω : when $r_0 \approx 0$, and for the ring of gas near equilibrium (ie $\Delta r \approx 0$ for all θ), by this equation, we should have :

$$-R(\omega - \omega_0)^2 = \frac{2}{R^2}$$

$$(\omega_0 - \omega)^2 = \frac{2}{R^3} \tag{3}$$

Then, applying Newton's 2nd law $a_r = f_r$ (physical forces) gives :

$$\begin{aligned} 0 \approx & \sum_{\sigma=\pm 1} (R + \sigma r_0 \cos\theta)(s_0^2 + 2\sigma r_0 R\cos\theta)^{-3/2} - \frac{1}{R^2} \\ & + \Delta r \sum_{\sigma=\pm 1} \left(-3(s_0^2 + 2\sigma r_0 R\cos\theta)^{-5/2}(R + \sigma r_0\cos\theta)^2 + (s_0^2 + 2\sigma r_0 R\cos\theta)^{-3/2} - \frac{1}{R^3} \right) \end{aligned}$$

We'll now use a-dimensional quantities by posing : $\delta := \dfrac{\Delta r}{R}$, $\epsilon := \dfrac{r_0}{R}$, which is equivalent to say that $R = 1$ if we redefine the lengths. For convenience, we write $\alpha := \cos(\theta)$. We then get :

$$\begin{aligned} 0 = & \sum_{\sigma=\pm 1} (1 + \sigma\epsilon\alpha)(1 + \epsilon^2 + 2\sigma\epsilon\alpha)^{-3/2} - 1 \\ & + \delta \sum_{\sigma=\pm 1} \left(-3(1 + \epsilon^2 + 2\sigma\epsilon\alpha)^{-5/2}(1 + \sigma\epsilon\alpha)^2 + (1 + \epsilon^2 + 2\sigma\epsilon\alpha)^{-3/2} - 1 \right) \end{aligned}$$

at order 1 in $\delta << 1 - \epsilon$, where we want to get $\delta(\epsilon, \alpha)$

We do now another Taylor expansion up to $\epsilon^2 + \delta\epsilon$, where we also suppose $\epsilon << 1$ (then $\delta << 1 - \epsilon$ iif $\delta << 1$).

$$\begin{aligned} 0 = & \sum_{\sigma=\pm 1} (1 + \sigma\epsilon\alpha) \left(1 - \frac{3}{2}\epsilon^2 - 3\sigma\epsilon\alpha + \frac{15}{2}\epsilon^2\alpha^2 + o(\epsilon^2) \right) - 1 \\ & + \delta \sum_{\sigma=\pm 1} (-3(1 - 5\sigma\epsilon\alpha + o(\epsilon))(1 + 2\sigma\epsilon\alpha + o(\epsilon)) + (1 - 3\sigma\epsilon\alpha + o(\epsilon)) - 1) \end{aligned}$$

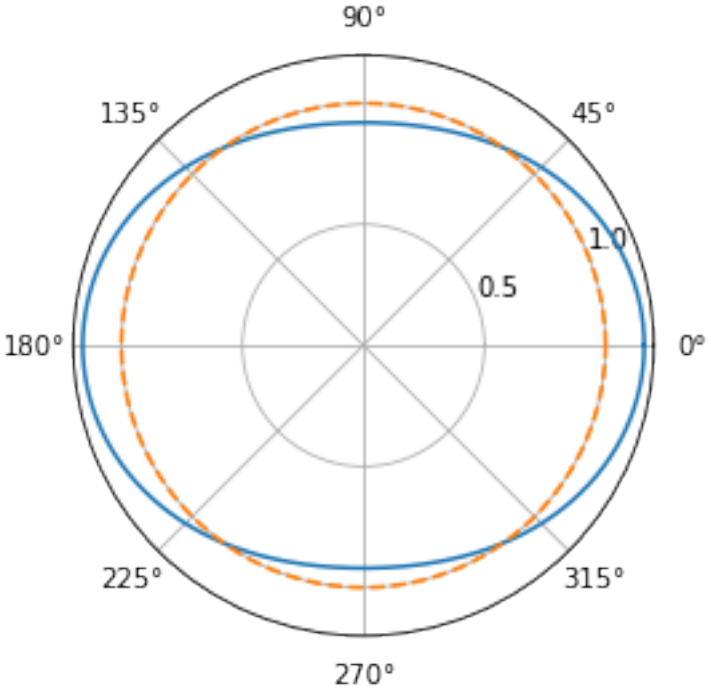

Figure 9: Tidal waves for $\delta, \epsilon << 1$

ie,

$$0 = -3\epsilon^2 + 9\alpha^2\epsilon^2 - 6\delta + o(\epsilon^2 + \delta\epsilon)$$

which gives,

$$\boxed{\delta(\epsilon, \cos\theta) \approx (3\cos(\theta)^2 - 1)\frac{\epsilon^2}{2}} \tag{4}$$

We get the following form for the chain of particles :

(Actually, this figure is drawn with a larger value for ϵ to be able to see the waves.)

Let compute the total mechanical energy of the ring of gas of total mass m, which is deformed by the tidal wave rotating at angular velocity ω in the non-rotating frame. For a particle of mass dm located at θ, its kinetic energy dKE is :

$$dKE = \frac{1}{2}dm\left(\frac{\partial\delta}{\partial t}\right)^2 \tag{5}$$

where now $\delta(\theta, t) = (3\cos(\theta - \omega t)^2 - 1)\frac{\epsilon^2}{2}$ We compute :

$$\begin{aligned}\frac{\partial\delta}{\partial t} &= -3\sin(\theta - \omega t)\cos(\theta - \omega t)\epsilon^2(-\omega) \\ &= \frac{3}{2}\sin(2(\theta - \omega t)\epsilon^2\omega\end{aligned}$$

We'll assume that the gas is uniformly distributed along the ring, so that we have $dm = \mu$ with $\mu = m$ (in our new unit of length where $R = 1$) Thus, the total kinetic energy carried by the gas is :

$$\begin{aligned} KE &\propto \epsilon^4\omega^2 \underbrace{\int_{-\pi}^{\pi} \sin^2(2\theta)\, d\theta}_{>0} \\ &\propto \epsilon^4\omega^2 \end{aligned}$$

where we've used that integrating over θ over one period is the same as integrating over $\theta - \omega t$.

On the other hand, the potential energy for one particle dPE comes only from the gravitational energy, thus : $dPE = -m\dfrac{1}{1+\delta} = -m + m\delta + o(\delta)$, so $dPE \approx m\delta$, when choosing the origin in energy. Then, the total potential energy is :

$$\begin{aligned} PE &\propto \epsilon^2 \underbrace{\int_{-\pi}^{\pi} (3 * \cos(\theta)^2 - 1)}_{>0} \\ &\propto \epsilon^2 \end{aligned}$$

But, $\omega^2 = 2 - \omega^2$, and by Kepler's third law applied the 2 stars on the same circular orbit $\omega^2 = \dfrac{1}{4\epsilon^3} >> 1$, thus $\omega^2 \propto \dfrac{1}{\epsilon^3}$. This gives $KE \propto \epsilon >> PE = O(\epsilon^2)$

Thus, the total energy δE_N ("N" stands for "Nutzbar" in German, which means "Usable") which can be extracted by the machine (which is supposed to be ideal, ie it extracts all the energy) is $\boxed{\delta E_N \propto \epsilon}$.

Then the process goes like this : during dt, the energy of the tidal wave created by the movement of the two stars gets extracted (instantaneously) by the machine into δE_N. Then, the ring of gas gets to its equilibrium state (ie $\delta(\theta) = 0$ for all θ). As the total energy of the system of {stars + gas + machine}, has to be conserved, there should be a decrease in the potential energy of the two stars, ie $d\epsilon < 0$, such that $\delta\left(\frac{1}{\epsilon}\right) \approx -\dfrac{1}{\epsilon^2}\dot{\epsilon}(t)\delta t = \delta E_N$, where δt is the time it takes to transfer the energy from the tidal wave to the machine, which is $\delta t = \dfrac{2\pi}{\omega} \propto \dfrac{1}{\epsilon^{3/2}}$ by relation 3, to the leading order. Considering $\delta\left(\frac{1}{\epsilon}\right)$ as an infinitesimal difference is just justified, since $\delta t << 1$. Thus, we get $\dot{\epsilon} \propto \dfrac{\epsilon^2 \delta E_N}{\delta t} \propto \epsilon^{3/2}$. Integrating this relation gives, $\epsilon^{-1/2} \propto t$, ie $\epsilon(t) = -\dfrac{C}{t^2}$ for $t > \tau$ where τ is some initial time and $C < 0$.

Then, we get the power transferred by the wave to the machine $P(t)$:

$$\begin{aligned} P(t) &= \frac{\delta E_N}{\delta t} \\ &\propto \frac{1}{\sqrt{\epsilon(t)}} \\ &\propto t \end{aligned}$$

The energy extracted by the machine $E_N(t)$ at time $t > \tau$ is then given by :

$$\begin{aligned} E_N(t) &\propto \int_\tau^t P(s)\, ds \\ &\propto (t-\tau)^2 \end{aligned}$$

Thus, $E_N(t)$ grows unbounded, this means we've got an arrow time, without specific initial conditions for the system.

Remark : if we look at the orbital momentum of the system {stars + gas}, it is not conserved as ω and ω_0 must have the same sign for $\epsilon \to 0$, as shown by (3). Therefore, some momentum goes also into the machine, which was supposed to be fixed at $r = 1$ and some fixed angle. We thus have to assume that a system in the machine keeps its orbital momentum equal to 0 (so that it doesn't rotate). For this, it has to bring some energy δE_R during δt, and we shall now verify that it doesn't compensate the energy extracted from the gas during the same amount of time.

Let's compute the amount δL_z of orbital momentum (along the z-axis - it is the only direction in which there is orbital momentum) transferred to the machine, it is :

$$\begin{aligned} \delta L_z &= -d(2M\omega_0\epsilon^2 + m\omega) \\ &\approx -d\left(M\sqrt{\epsilon} + \frac{m}{2}\epsilon^{-3/2}\right) \\ \delta L_z &\approx -\frac{3m}{4}\epsilon^{-5/2}\dot{\epsilon}\delta t \end{aligned}$$

We'll find that δE_N and δE_R are of the same order (in t^2), so we have to compute the

constants. If we look again at the computations of δE_N, we get :

$$\begin{aligned}\delta E_N &= \frac{1}{2}.\frac{9}{4}.\frac{1}{2}(2\pi)\epsilon^4\omega^2 \\ &\approx \frac{9m}{32}\epsilon\end{aligned}$$

$$\begin{aligned}\frac{-M}{2\epsilon^2}\dot{\epsilon}\delta t = \delta E_N &= \frac{9m}{32}\epsilon \\ \dot{\epsilon}\delta t &= -\frac{9m}{16M}\epsilon^3\end{aligned}$$

Thus,

$$\delta L_z = \frac{3^3 m^2}{2^6 M}\epsilon^{1/2}$$

Compared to the size of the ring of gas, we assume that the size of the machine is much smaller, so we can approximate it as a particle point of mass M_E ("E" stands for experiment). The rotational kinetic energy associated to it is then given by :

$$\delta E_R = \frac{m}{2}\omega_E^2$$

where ω_E is the angular velocity of the machine of mass M_E that comes from the momentum brought by the system stars + gas and which is set to be 0 afterwards (in an amount of time less than δt considered to be infinitesimal) by some system in the machine. One has :

$$\omega_E = \frac{\delta L_z}{M_E}$$

Thus, one gets :

$$\begin{aligned}\delta E_R &= \frac{1}{2M_E}\left(\frac{3}{4}\right)^6 \frac{m^4}{M^2}\epsilon \\ &= \frac{9m^4}{32M_E M^2}\left(\frac{3}{4}\right)^4 \epsilon\end{aligned}$$

Finally, we get :

$$\frac{\delta E_R}{\delta E_N} \approx \frac{m^3}{M_E M^2}\left(\frac{3}{4}\right)^4 << 1$$

Does it sound plausible ? Suppose m corresponds to the mass of the asteroid belt in the Solar system, ie $m \approx 10^{21}$ kg, M the mass of the Sun, ie $M \approx 10^{30}$ kg, then : $(m/M)^2(3/4)^2 \approx 10^{-19}$

To have $\frac{\delta E_R}{\delta E_N} < 0.01$ for example, one should take $M_E > 10^{-17}m \approx 10^3$ kg which sounds physically plausible.

In conclusion, if we make this last assumption concerning the masses, the energy that the machine should bring to keep itself at rest doesn't compensate the energy extracted from the tidal wave, and we still get an unbounded increase of the energy gotten from the machine, ie there is an arrow time in the direction of this increase.

7 Conclusion

Through this work, we saw that explaining thermodynamical behavior with the laws of mechanics is a difficult problem and leads to profound philosophical questions, such as knowing why time flows only towards the future, and not towards the past.
We explored some ideas of Boltzmann's attempt to build a bridge between Newton's reversible laws and the irreversible 2nd law of thermodynamics, by using typicality arguments.
Doing this, the question of the existence of non-equilibrium states (like the current macro state of our universe) arises and one may answer this question by stating that our universe was set at the Big Bang into specific initial conditions.
In this work, we've tried to build models where we would not need to specify such peculiar initial conditions to get an arrow of time, by looking at gravitational systems. Finally, we dealt with a binary star system surrounded by a rotating ring of gas, where a physical reasoning suggests that an arrow of time could appear in this system (which corresponds in this case to an unbounded increase of the extracted kinetic energy), whatever the initial conditions might be.
The next step for a future investigation would be to provide a mathematical proof of this, but it is much more difficult to derive a rigorous statement, a challenge that my tutor and his PhD student Aaron Schaal are working on.

A Python code for the numerical experiment

```
import numpy as np
import matplotlib.pyplot as plt

"""
Create Your Own N-body Simulation (With Python)
Philip Mocz (2020) Princeton Univeristy, @PMocz
Simulate orbits of stars interacting due to gravity
Code calculates pairwise forces according to Newton's Law of Gravity
"""
L = 1000  #length of the torus in one direction

def abstand(dx):  #computes the distance (in one direction) between two stars on the torus  ; "abstand" means "distance"
    dx_inv = dx-L*np.sign(dx)
    abstand, abstand_inv = abs(dx), abs(dx_inv)

    if abstand < abstand_inv :
        return dx
    else :
        return dx_inv

abstand_vec = np.vectorize(abstand)

def getAcc( pos, mass, G, softening ):
    """
    Calculate the acceleration on each particle due to Newton's Law
    pos  is an N x 3 matrix of positions
    mass is an N x 1 vector of masses
    G is Newton's Gravitational constant
    softening is the softening length
    a is N x 3 matrix of accelerations
    """
    # positions r = [x,y,z] for all particles
    x = pos[:,0:1] % L
    y = pos[:,1:2] % L
    z = pos[:,2:3] % L

    # matrix that stores all pairwise particle separations: r_j - r_i
    dx = abstand_vec(x.T - x)
    dy = abstand_vec(y.T - y)
    dz = abstand_vec(z.T - z)

    # matrix that stores 1/r^3 for all particle pairwise particle separations
    inv_r3 = (dx**2 + dy**2 + dz**2 + softening**2)
    inv_r3[inv_r3>0] = inv_r3[inv_r3>0]**(-1.5)

    ax = G * (dx * inv_r3) @ mass
    ay = G * (dy * inv_r3) @ mass
```

```
    az = G * (dz * inv_r3) @ mass

    # pack together the acceleration components
    a = np.hstack((ax,ay,az))

    return a

def getEnergy( pos, vel, mass, G ):
    """
    Get kinetic energy (KE) and potential energy (PE) of simulation
    pos is N x 3 matrix of positions
    vel is N x 3 matrix of velocities
    mass is an N x 1 vector of masses
    G is Newton's Gravitational constant
    KE is the kinetic energy of the system
    PE is the potential energy of the system
    """
    # Kinetic Energy:
    KE = 0.5 * np.sum(np.sum( mass * vel**2 ))


    # Potential Energy:

    # positions r = [x,y,z] for all particles
    x = pos[:,0:1] % L
    y = pos[:,1:2] % L
    z = pos[:,2:3] % L

    # matrix that stores all pairwise particle separations: r_j - r_i
    dx = abstand_vec(x.T - x)
    dy = abstand_vec(y.T - y)
    dz = abstand_vec(z.T - z)

    # matrix that stores 1/r for all particle pairwise particle separations
    inv_r = np.sqrt(dx**2 + dy**2 + dz**2)
    inv_r[inv_r>0] = 1.0/inv_r[inv_r>0]

    # sum over upper triangle, to count each interaction only once
    PE = G * np.sum(np.sum(np.triu(-(mass*mass.T)*inv_r,1)))

    return KE, PE;


def main():
    """ N-body simulation """

    # Simulation parameters

    N         = 10    # Number of particles
    t         = 0      # current time of the simulation
    tEnd      = 50   # time at which simulation ends
```

```
    dt         = 0.000001   # timestep
    softening = 0.0001    # softening length
    G          = 10**-3    # Newton's Gravitational Constant
    plotRealTime = False # switch on for plotting as the simulation goes
    along

    # Generate Initial Conditions
    np.random.seed(17)            # set the random number generator seed

    mass = 20.0*np.ones((N,1))/N  # total mass of particles is 20
    mass[0][0] = 10**6 * mass[0][0]

    pos  = (L * np.random.randn(N,3)) % L  # randomly selected positions
    and velocities
    vel  = 1/L * np.random.randn(N,3)

    # Convert to Center-of-Mass frame
    vel -= np.mean(mass * vel,0) / np.mean(mass)

    # calculate initial gravitational accelerations
    acc = getAcc( pos, mass, G, softening )

    # calculate initial energy of system
    KE, PE  = getEnergy( pos, vel, mass, G )

    # number of timesteps
    Nt = int(np.ceil(tEnd/dt))

    # save energies, particle orbits for plotting trails
    pos_save = np.zeros((N,3,Nt+1))
    pos_save[:,:,0] = pos
    KE_save = np.zeros(Nt+1)
    KE_save[0] = KE
    PE_save = np.zeros(Nt+1)
    PE_save[0] = PE
    t_all = np.arange(Nt+1)*dt

    # prep figure
    fig = plt.figure(figsize=(4,5), dpi=80)
    grid = plt.GridSpec(3, 1, wspace=0.0, hspace=0.3)
    ax1 = plt.subplot(grid[0:2,0])
    ax2 = plt.subplot(grid[2,0])

    # Simulation Main Loop
    for i in range(Nt):
        # (1/2) kick
        vel += acc * dt/2.0

        # drift
        pos += vel * dt
        pos = pos % L
```

```
148
149         # update accelerations
150         acc = getAcc( pos, mass, G, softening )
151
152         # (1/2) kick
153         vel += acc * dt/2.0
154
155         # update time
156         t += dt
157
158         # get energy of system
159         KE, PE  = getEnergy( pos, vel, mass, G )
160
161         # save energies, positions for plotting trail
162         pos_save[:,:,i+1] = pos
163         KE_save[i+1] = KE
164         PE_save[i+1] = PE
165
166     # plot in real time
167     i = Nt-1
168     if plotRealTime or (i == Nt-1):
169         plt.sca(ax1)
170         plt.cla()
171         xx = pos_save[:,0,max(i-50,0):i+1]
172         yy = pos_save[:,1,max(i-50,0):i+1]
173         plt.scatter(xx,yy,s=1,color=[.7,.7,1])
174         plt.scatter(pos[:,0],pos[:,1],s=10,color='blue')
175         ax1.set(xlim=(0, L), ylim=(0, L))
176         ax1.set_aspect('equal', 'box')
177         #ax1.set_xticks([-2,-1,0,1,2])
178         #ax1.set_yticks([-2,-1,0,1,2])
179
180         plt.sca(ax2)
181         plt.cla()
182         plt.scatter(t_all,KE_save,color='red',s=1,label='KE' if i == Nt-1
    else "")
183         plt.scatter(t_all,PE_save,color='blue',s=1,label='PE' if i == Nt-1
    else "")
184         plt.scatter(t_all,KE_save+PE_save,color='black',s=1,label='Etot' if
     i == Nt-1 else "")
185         plt.xlabel('time')
186         plt.ylabel('energy')
187         ax2.legend(loc='upper right')
188         ax2.set(xlim=(0, tEnd))
189
190
191     # Save figure
192     #plt.savefig('nbody.png',dpi=240)
193     plt.show()
194
195     return pos, PE_save, KE_save + PE_save, mass
```

```
196
197 if __name__== "__main__":
198
199     pos, PE_save, H, mass = main()
```

B Python code to plot the tidal waves

```
1  import numpy as np
2  import matplotlib.pyplot as plt
3  from math import *
4
5  N = 1000
6  eps = 0.4
7
8
9  def radius(theta, eps):
10     return 1 + (3*np.cos(theta)**2 -1) * eps ** 2 /2
11
12 theta = np.linspace(0, (2 * np.pi), N)
13 r = np.vectorize(radius)(theta, eps)
14
15 fig, ax = plt.subplots(subplot_kw={'projection': 'polar'})
16 ax.plot(theta, r)
17 ax.plot(theta, [1 for k in range(N)], '--')
18 ax.set_rmax(1.2)
19 ax.set_rticks([0.5, 1])
20 ax.grid(True)
21
22 plt.show()
```

References

[1] Roderich Tumulka. Lecture notes on mathematical statistical physics. Summer Semester 2019.

[2] Jean Bricmont. SCIENCE OF CHAOS OR CHAOS IN SCIENCE? *Annals of the New York Academy of Sciences*, 775(1):131–175, 06 1995.

[3] V. A. Antonov. Most probable phase distribution in spherical star systems and conditions for its existence. In J. Goodman and P. Hut, editors, *Dynamics of Star Clusters*, volume 113, page 525, January 1985.

[4] Philip Mocz. Create your own n-body simulation (with python), https://github.com/pmocz/nbody-python (accessed: 04.07.2022).