

The Seven Wonders of the World

Lecture notes for ING 175

Version 0.1, updated 1 March 2024



CC BY SA Licence

Typeset with L^AT_EX using 12 pt Palatino and Optima fonts

Cover image: the “Pale Blue Dot” image of the Earth taken by Voyager 1,
from right outside Pluto’s orbit

<https://science.nasa.gov/resource/voyager-1s-pale-blue-dot/>

Introduction

The loss implied in such an acquisition can be estimated only by those who have been compelled to unlearn a science that they might at length begin to learn it.

J. C. Maxwell 1878

Until some decades ago, the 18th-century physical notions typically taught in introductory Bachelor physics courses were enough to prepare an engineer for future specializations and jobs. Students who wanted to venture into modern theories, such as Relativity, were required to **re-learn** some of the most important physical notions – *Energy*, mass, time, entropy above all – which in these theories are quite different from the 18th-century ones. But at that time the modern theories still mostly had only theoretical, not practical, importance. So the re-learning efforts of the curious students could perhaps be justified.

That situation has changed today. Modern theories are an essential part of many everyday technologies, like nuclear reactors and the [Global Positioning System](#)¹ (see the entertaining discussion in Taylor & Wheeler 2000 Project A, and also Petit & Wolf 2005; Fliegel & DiEsposti 1996; Ashby 2002; Müller et al. 2008); and they are required for developing new technological possibilities, from [quantum computers](#)² to [solar sails](#)³ (see for instance the [CubeSail](#)⁴ mission). An engineering student (including communication and data engineering) may likely end up in a job that requires an understanding of modern physical notions. The diffusion of [large language models](#)⁵ will moreover require future engineers who actually **understand** those physical notions, not little monkeys who have been trained to manipulate some equations and to throw some technobabble around. Automated large language models are faster, cheaper, and more precise in doing the

"Numerous relativistic issues and effects play roles in the global positioning system, on which millions of drivers, hikers, sailors, and pilots depend to find out where they are. The GPS system is, in effect, a realization of Einstein's view of space and time. Indeed, the system cannot function properly without taking account of fundamental relativistic principles." Ashby 2002

latter kind of monkey activities. So why should one hire a human to do the same?

While moving from the older to the newer notions often requires re-learning efforts and conceptual re-orientations, the move in the opposite direction is less demanding. The modern physical notions are more encompassing than their 18th-century parents. Their understanding leads to an understanding of their older counterparts as approximate and special cases. Students, moreover, have often been hearing quite early from mass media about the new notions; for instance about the equivalence of mass and energy. Owing to this early exposure, students sometimes ask very intelligent and deep questions – such as “*should the mass of the body be included in its internal energy?*” – when exposed to the old notions.

It is therefore high time that introductory Bachelor physics courses be based on modern physical notions. Students should not be required to waste time and mental effort to learn something that they must unlearn and re-learn, only because of academia’s and teachers’ inertia.

Some teachers say “it would be too difficult for students to understand modern ideas, because they are too familiar with the old ones. This is why we need to teach the old and slightly incorrect ideas first, and correct them later”. I think that this kind of reasoning is scientifically unacceptable and leads to a vicious circle. Students are unfamiliar with new notions only because they were raised by a generation who was taught old ones. New notions become familiar only after a couple of generations learn them at an early stage. This is obvious if you consider that notions such as “energy”, “electromagnetic field”, “vector” are very familiar today, but were absolutely *unfamiliar* a couple centuries ago. If we had always taught what’s familiar, then we would still be teaching about the elements *air, earth, water, fire*, and that the Sun revolves around the Earth. Arguments in favour of teaching old notions are for the most part just pretexts for laziness.

The present lecture notes are an experiment and an attempt to introduce classical mechanics and thermodynamics from modern physical notions. The core equations remain the same, but the students should have a broader conception of their meaning and of the symbols that appear in them.

URLs for chapter *Introduction*

1. <https://www.gps.gov>
2. <https://www.ibm.com/topics/quantum-computing>
3. <https://www.planetary.org/sci-tech/lightsail>
4. <https://www.cubesail.us>
5. <https://www.ibm.com/topics/large-language-models>

Chapter 0

Guide to the text

Maths prerequisites

- Working familiarity with algebra, its operations, and their properties, including square roots; knowledge of exponentials and logarithms is useful.
- Working familiarity with solving equations and inequalities, linear and non-linear.
- Working familiarity with the study of functions of one real variable.
- Working familiarity with derivatives.
- Understanding of what an integral is, even if you won't be required to solve integrals.
- Working familiarity with vector calculus.
- Some familiarity with functions of many variables.
- Understanding of what partial derivatives are.

Physics prerequisites

Just some vague reminiscences of secondary/high-school physics should be enough.

It can be beneficial if you are familiar with basic physics notions such as *velocity*, *mass*, *force*, and similar ones; and if you vaguely remember about Newton's laws.

Structure of this text

👉 Graphical devices

The text includes the following graphical devices:

- Important notions and definitions are also given in **boldface**.
- Important-notion boxes:

📘 Something

This is a definition or explanation of Something.

- Warnings and important points that require careful thinking:

❗ Something

Something you must be careful about.

- Exercises:

💡 Exercise 0.1

This isn't an exercise

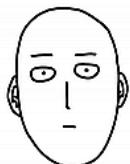
- Discussions and connections with more advanced physics:

Почем How things really are in quantum physics

Just for your curiosity: you don't need to remember any of this.

👉 Side pictures and quotes

Pictures, graphs, or quotes related to the material are displayed on the right.



This is an image of Saitama, which actually has nothing to do with the text on the left.

👉 Hyperlinks and bibliography

Some words are hyperlinks, like this one about [One Punch Man¹](#); you also recognize them because they have a little footnote number. The links' URLs are listed at the end of each chapter, just in case you're reading a printed copy and wonder what the link was.

The text gives bibliographic references, like “Einstein 1905a”, to scientific literature. The references are listed in the final Bibliography on page 125.

These references are not part of the course and don't need to be “studied”. They are given for two reasons:

- For your own curiosity.
- To back up what's written in the text. In science you should not believe something just because you've read it somewhere. You should, as much as possible, go and check *for yourself how the logic behind the statement is proved and what is the experimental evidence behind the statement*.

👉 Notation and terminology

Mathematical notation, as well as notation for physical dimensions, strictly follows the standards given by the International System of Units (SI)², listed for example in ISO 2009 and ISO 2019.

“Believe nothing, O monks, merely because you have been told it, or because it is traditional, or because you yourselves have imagined it. Do not believe what your teacher tells you merely out of respect for the teacher.”

(attributed to Gautama Buddha)

URLs for chapter 0

1. <https://onepunchman.fandom.com>
2. <https://www.nist.gov/pml/special-publication-811>

Chapter 1

Physics?

If you think about it, many things we ordinarily do every day are some sort of magic. Think of how you can instantaneously see and speak with a person living on another continent, in real time, using just a small widget in the palm of your hand. Think of how you can instantaneously see where you are on the Earth, using the same widget. Think of how fast you can go to another country, by flying in a huge metal thing. Think of how you can command and interact with a purely fictitious animated world when you play on your computer. The list can go on forever. Other things are luckily less ordinary, but still inspire a lot of awe: think of the devastating power unleashed by something roughly as small as a tennis ball, in an atomic bomb.

We can do these astonishing things thanks to our understanding of how the world works. That's Physics.

Many things can be said and have been said about science and physics. Rather than repeating what's been already written in many excellent books, I invite you to take a break here and go read their introductions. Choose as you please; compare what they say; don't limit yourself to popular books.

.....

1.1 Several possible formalisms or “languages”

Physics can be expressed and written from wildly different points of view, using wildly different principles. Let's call these “different physics languages”; a more technical name is “physics formalisms”. One may approach a physics phenomenon or problem in terms of *Lagrangeans*, or

Hamiltonians, or *fibre bundles*, or *categories*, or *action principles*, or many other formalisms. These formalisms or languages are not completely separated; we know how to translate among them. In "doing" physics, one may jump among formalisms, because some ideas may be easier to express, or some results easier to find, in one formalism than another. No matter which physics formalism you choose, the results and the concrete applications are still the same. The choice is to a great extent subjective, based on your aesthetic tastes. You see that in "doing" physics you can express your personality and put your own artistic touch; this is why it's such a cool subject (and other scientific subjects are like this too).

In these notes I'm choosing one particular formalism: the one that for me is the most easily *visualizable*; because I believe that visualization can be beneficial in learning new things. Or maybe I'm choosing it just because I like it best. I encourage you to explore how the physics you've learned is expressed in other physics formalisms; maybe you'll like another physics formalism better.

The formalism we'll be using might be called "field theory". Roughly speaking it takes as starting point the ideas of space and time, or better spacetime, in which there are different kinds of "stuff". It expresses the regularity and patterns that we observe in physical phenomena as "budgets" about the different kinds of stuff, and of relations between these kinds. Please don't take the description just given too literally; it's just meant to give you a very vague idea of the field-theoretical viewpoint.

It goes without saying that all these "physics languages" are to a great extent mathematical.

One reason is that numbers allow us to convey information in a concise and precise way. Imagine you have to tell someone, who doesn't know Bergen, where in Bergen you are right now, to within 10 m. You can do that with a description, "... and there's a building called so-and-so which looks like so-and-so...", which would be lengthy and tricky. Or you can just give two numbers: latitude and longitude:

60.369 40, 5.3518 .

And in these two numbers all digits are important; for instance, the latitude is not 60.369 47.

But the most important reason is that mathematics allows us to describe and follow the patterns and variety of physical phenomena in a greatly concise and precise way. And to develop their relationships in a rigorous

$$\delta \int L dt = 0 \quad L = \frac{1}{2} mv^2$$

$$\mathbf{F} = \frac{d}{dt} m \mathbf{v} \quad \mathbf{F} = 0$$

Example of two different formalisms (red, blue) expressing the same physical phenomenon.

"this grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics"

Galilei 1623

way. All our present technology would have been impossible to discover, and would be impossible to realize, without the mathematical language of physics.

I invite you again to read what many good texts say about the relationship between physics and mathematics. No point repeating here what is said better elsewhere.

1.2 Quantities: primitive and derived

One topic must be briefly discussed because it's important for understanding the notes that follow. It's the distinction between *primitive* and *derived* quantities.

I shall assume that you already know what a **physical quantity** is. Examples are: position, duration, velocity, pressure, energy, temperature.

A **derived quantity** is one that is defined in terms of other quantities. For example, velocity v (more precisely: average velocity) is defined as the ratio between a distance d (a vector) and a time duration t :

$$v := \frac{d}{t}$$

where the symbol “:=” means “is defined as” or “is defined by”. This means that in principle we could avoid using the word “velocity” and the symbol “ v ” altogether, and instead always speak about distance and duration, using their symbols. It would lead to very long sentences and formulae and would be extremely inconvenient, but it could be done. The definition of a derived quantity often tells us how that quantity can be measured.

A derived quantity is defined in terms of other quantities, and these may in turn be derived quantities, that is, defined in terms of still other quantities, and so on. But at some point this chain of definitions must come to an end, otherwise we would go around in circles.

A **primitive quantity** is one that we do not define in terms of other quantities. Primitive quantities are the building blocks from which we define all others. That they are not defined in terms of others doesn't mean that we cannot try to explain them. But such explanations must be taken as informal and heuristic. Primitive quantities are often explained through metaphors and by appealing to intuition. You must always be wary of such explanations, because they may fail you spectacularly in some situations.

Often we have a choice about which quantities should be primitive and which should be derived. For instance, *energy* can be defined, in a

“There is nothing that can be said by mathematical symbols and relations which cannot also be said by words. The converse, however, is false. Much that can be and is said by words cannot successfully be put into equations, because it is nonsense.”

Truesdell 1966



“you have to be in some framework that you allow something to be true. Otherwise you're perpetually asking ‘why’”. (see video¹).

somewhat complicated way, in terms of quantities like *work* and *heat*, which would then need to be taken as primitive. Or we can take *energy* as primitive, and define *work* and *heat* in terms of it. This second choice can be more convenient to develop a physical theory. It often happens that a quantity is very convenient for building a theory, if used as primitive; but difficult to understand intuitively. Vice versa, a quantity can be very intuitive but lead to a complicated theory. Among quantities which we'll take as primitive are: *time*, *space* and *length*, *matter*, *energy*, *momentum*, *entropy*, *temperature*, and several others. All will be discussed soon.

1.3 Physical dimensions and units

Measurement is the process by which we determine the value of a physical quantity. Measurements can be extremely complex, and can extremely different even if they are about the same quantity. Consider the ways we can measure the mass of a football, compared to the ways we can measure the mass of the Sun.

To each quantity we associate a **physical dimension**. The term ‘dimension’ here has nothing to do with physical extension, as in “the dimensions of this box”; be careful not to confuse the two. Usually it’s clear which one is meant from the context. Physical dimensions help us avoid making operations that don’t make sense with some quantities. For example, it doesn’t make sense to sum up the volume of a body of water with its temperature; and indeed the volume has dimension *length*³, whereas temperature has dimension *temperature*, and quantities with different dimensions cannot be added up.

With each physical dimension we can associate a measurement **unit**, which expresses a basic standard for comparing the measurement results of similar quantities. For example, we could use the *minute* or the *second* as units to measure *time*.

One can choose a basic set of physical dimensions from which to define all others, and for these a set of standard units. Here we shall follow the [International System of Units \(SI\)](#)² (see also [NIST Special Publication 811](#)³).

The topics of measurement and physical dimensions, which are studied in *metrology* and in *dimensional analysis*, could occupy an entire course by themselves. I shall assume that you already know their basics notions and that you read about the SI.

The measurement of some physical quantities consists in just one number with associated physical dimension; we shall call such quantity a

scalar. The measurement of other physical quantities consists instead in a triplet of numbers with associated physical dimension; we shall call such quantity a **vector**.

! What's scalar or vector depends on the theory

Scalar and *vector* have very specific and slightly different meanings in different theories, so don't take the definitions used here as universal. For example, in these notes and in Newtonian mechanics we call *energy density* a scalar, but in general relativity it cannot be called a scalar.

1.4 Informal tips about units and maths

Importance of units

Units are very important and must always be written for several reasons.

First, a number without units doesn't tell us anything. If I tell you "the place is at a distance 100 from here", you have no idea how far the place is. "100" what? 100 metres? 100 kilometres? These are completely different distances.

Second, units give us useful information about mathematical objects and their physical relationships and measurement. If you see the expression "3 m/s", then there's a strong possibility that that's a velocity. If you see the expression "5 J/m²", then you have a hint that it could be measured by measuring an energy and then dividing by an area.

Third, because of the previous reason, keeping track of units often allows us to quickly catch errors in solving a physical problem.

Variables and units

When a physical quantity is denoted by a symbol or variable, keep in mind that a unit is "contained" in the symbol, so to speak. For example if the variable t denotes a time, then it includes some time unit, say seconds. This becomes apparent when we write the value of the symbol, for instance " $t = 120 \text{ s}$ ". The unit is not predetermined, but it must correspond to the dimension of that quantity. We could for instance write " $t = 2 \text{ min}$ " instead; the two expressions are completely equivalent.

This fact must be kept in mind when combining symbols. For example, if d is a distance and t is a time, then writing $v = d/t$ tells us that v is a

velocity, and it has appropriate units that come from d and t , for instance m/s.

Units otherwise behave just like *literal constants* for all mathematical purposes, just like the letter ‘ a ’ in the expression ‘ $a x$ ’. This is why they can, for instance, be simplified:

$$3 \text{ mol/s} \cdot 5 \text{ s} = 3 \frac{\text{mol}}{\text{s}} \cdot 5 \text{ s} = 15 \text{ mol} .$$

Mathematical functions and units

Particular care must be taken with trigonometric and exponential functions, such as $\sin()$, $\cos()$, $\tan()$, $\exp()$, $\log()$; **these functions only admit a dimensionless argument** (which for the trigonometric ones corresponds to *radians*). So there cannot be units like ‘s’ or ‘m’ within these functions: we must make sure that any units present within cancel out.

This makes sense, because we wouldn’t know how to interpret the argument otherwise. Suppose you read “ $\cos(60 \text{ s})$ ” somewhere: how much is that? If we say “just discard the unit”, we would have

$$\cos(60 \text{ s}) \stackrel{?}{=} \cos(60) \approx -0.95$$

but wait: $60 \text{ s} \equiv 1 \text{ min}$, so we could equivalently write “ $\cos(1 \text{ min})$ ”. Then, according to the hypothetical rule “just discard the unit”, we would have

$$\cos(60 \text{ s}) \equiv \cos(1 \text{ min}) \stackrel{?}{=} \cos(1) \approx +0.54$$

a completely different result!

For this reason an expression like ‘ $\cos(t)$ ’, with t denoting time, doesn’t make sense: there’s a time unit in the argument of $\cos()$. If we want to express an oscillation with time, we must write instead something like

$$\cos\left(\frac{t}{T}\right)$$

where T is the period of the oscillation, a symbol which also includes a time unit, which simplifies with the one in t . If the period of the oscillation is $T = 1 \text{ s}$ then we can also simply write

$$\cos(t/\text{s})$$

This expression is now unambiguous: suppose that $t = 60 \text{ s} \equiv 1 \text{ min}$, then

$$\begin{aligned} \cos(t/\text{s}) &= \cos(60 \text{ s}/\text{s}) = \cos(60) \approx -0.95 \\ &= \cos(1 \text{ min}/\text{s}) = \cos(1 \cdot 60 \text{ s}/\text{s}) = \cos(60) \approx -0.95 \end{aligned}$$

Also remember that **the results of trigonometric and exponential functions are dimensionless numbers**, so an expression like ‘ $3 \cos(\dots)$ ’ denotes a pure number, with no units. If you want to express that the result is a length, the appropriate units must appear. We can for instance write

$$L \cos(\dots)$$

where L is a length, and therefore includes some kind of length unit such as ‘m’. If this length is, say, $L = 2 \text{ m}$ we can also simply write

$$2 \cos(\dots) \text{ m}$$

Units and derivatives

When we follow the rules above, all other mathematical operations automatically take care of everything. The derivative, for instance, is calculated in the usual way, treating any visible units as literal constants. Let’s see a concrete example. This expression

$$x(t) = 2 \cos(t/\text{s}) \text{ m}$$

says that the position of some object oscillates with time, between the values -2 m and $+2 \text{ m}$. When $t = 0 \text{ s}$, the position is $x = +2 \text{ m}$. The position $x = -2 \text{ m}$ is reached when the argument of $\cos()$ is π , that is

$$t/\text{s} = \pi \quad \Rightarrow \quad t \approx 3.14 \text{ s}.$$

The velocity of the object (see § 2.4) is given by the derivative of this expression with respect to t . Let’s calculate it treating all unit symbols as literal constants:

$$\frac{dx(t)}{dt} = \frac{d}{dt} \left(2 \cos(t/\text{s}) \text{ m} \right) = 2 \underbrace{\left[-\sin(t/\text{s}) \cdot \frac{1}{\text{s}} \right]}_{\text{chain rule}} \text{ m} = -2 \sin(t/\text{s}) \text{ m/s}$$

and you see that the correct units for velocity have automatically appeared.

1.5 What is “fundamental” physics?

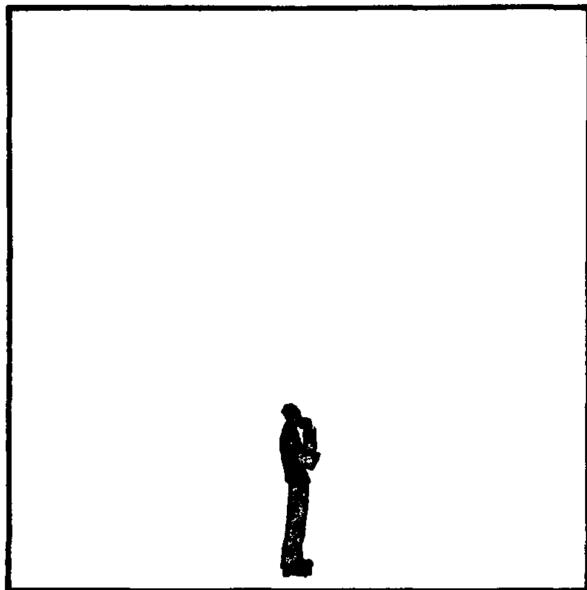
But what’s the “ultimate” goal of physics? What’s “fundamental” physics? The answer to this question is again subjective – also in this case physics lets you express your proclivities and personality. In the history of physics one can probably identify two main conceptions of “fundamental” physics.

For some physicists it is about finding the ultimate building blocks, so that one day we can say “... and these are the constituents, and they obey these equations”. The history of physics seems to show that this goal is overturned every few generations. And yet every generation says “Now we almost have the complete picture – it’s right behind the corner. It’s true that previous generations thought they almost had it, and turned out to be wrong. But *this time* is different, this time we have the real deal!”. The theoretical and particle physicist [Geffrey Chew⁴](#) depicted this situation as in fig. 1.1. For this reason some physicists are a little sceptical about this goal; maybe it’s a never-ending structure, with surprises at every deeper look.

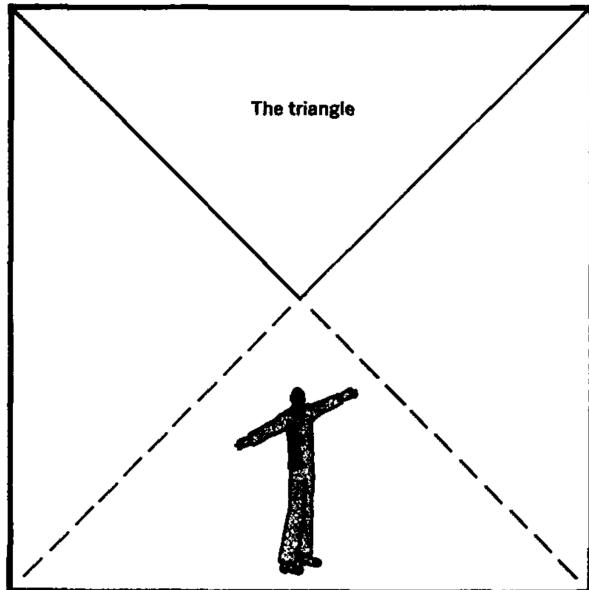
So for other physicists fundamental physics is about finding some point of view or mathematical structure that is rich enough to make useful predictions, and yet flexible enough to accommodate any new patterns or objects that we might discover. In a manner of speaking, it is about finding “patterns of patterns” or “laws about physical laws”.

The two conceptions above are not mutually exclusive, and both are always pursued, even if time-changing fashions may emphasize the one or the other.

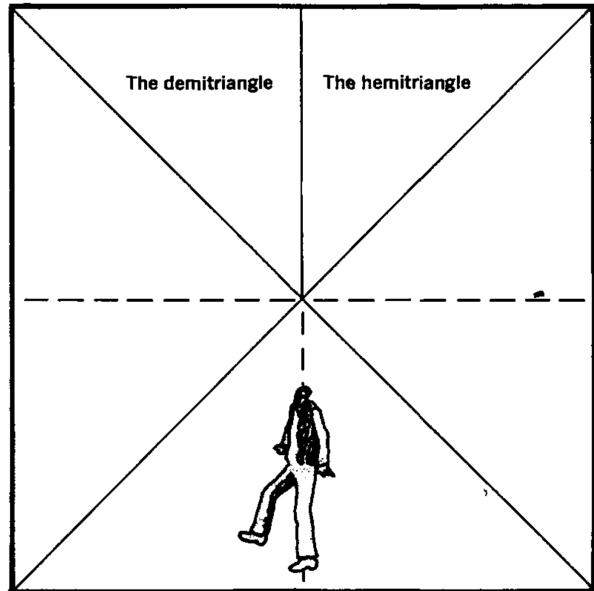
In these notes we take a point of view slightly closer to the second conception. This will also be reflected in the main division between physical laws that we’ll draw in ch. 5.



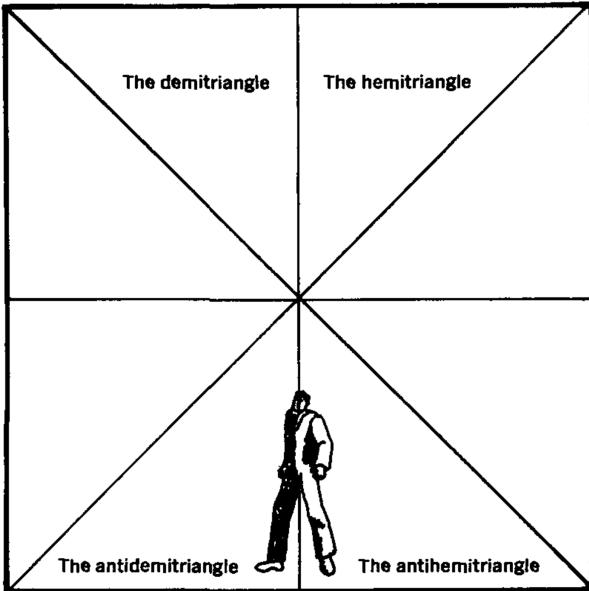
10 000 BC. The inhabitants of the paper square have no conception of the true nature of the universe they inhabit.



1900 AD. Physicists of the square discover a basic subdivision of their universe. They call it the "triangle" and consider it to be the fundamental building block of the universe.

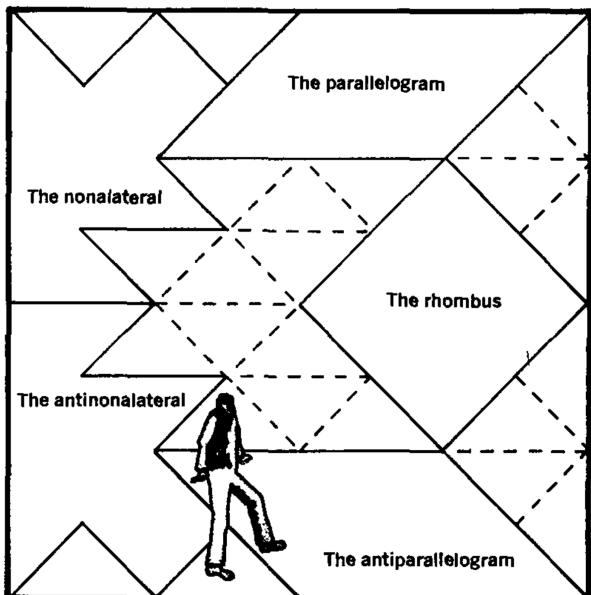


1930 AD. Physicists discover that the triangle can be split. Its parts are termed the "hemitriangle" and the "demitriangle." These are thought to be the fundamental building blocks of the universe.

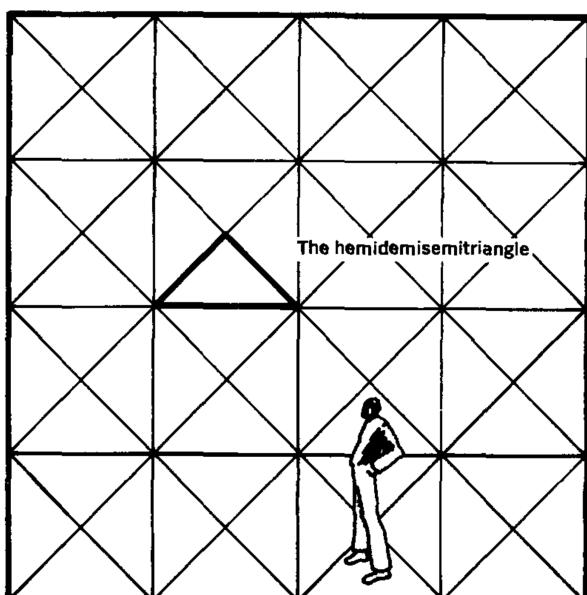


1950 AD. Mirror images of the hemitriangle and the demitriangle are discovered. These are termed "antihemitriangle" and "antidemitriangle."

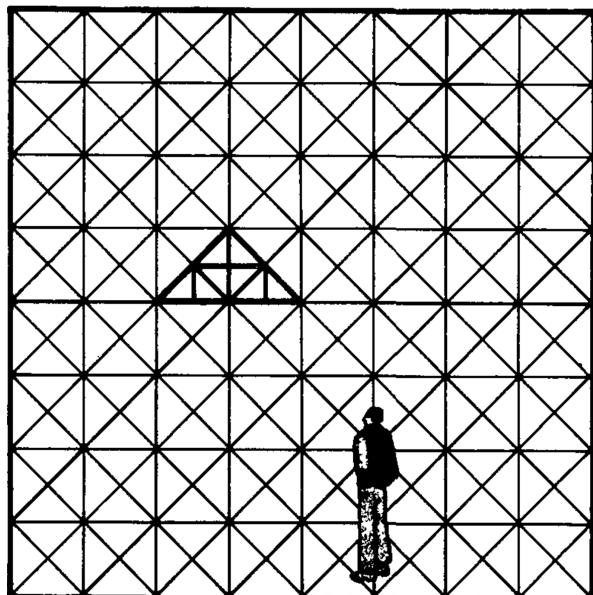
Figure 1.1 (Continues on p. 17) *The progress of "fundamental" physics*, from Chew 1970 as reproduced in Truesdell 1987



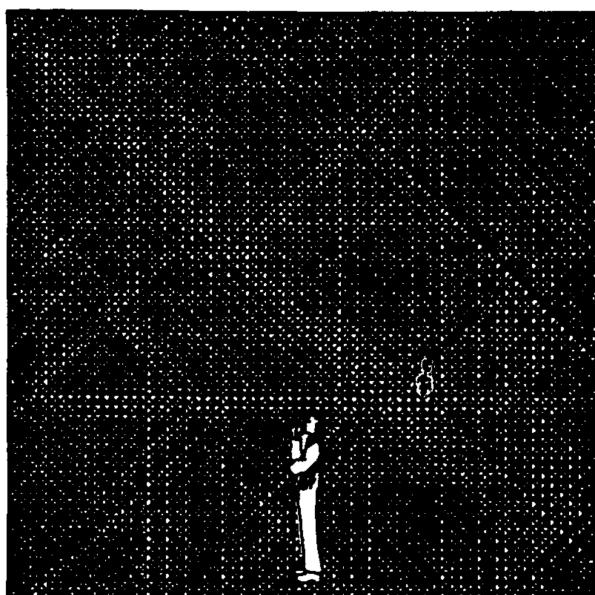
1960 AD. Physicists' conception of their universe is further clouded by new discoveries: the rhombus, the parallelogram, the antiparallelogram, the nonalateral and many others. It is unclear what these discoveries signify.



1970 AD. A new configuration, the "hemidemisemitriangle," is hypothesized, out of which all known configurations of the universe can be constructed. The hemidemisemitriangle is thought to be the fundamental building block of the universe.



1975 AD. The hemidemisemitriangle is discovered. The following year the hemidemisemitriangle is split.



2000 AD. The inhabitants of this paper square have no conception of the true nature of the universe they inhabit.

URLs for chapter 1

1. <https://www.youtube.com/watch?v=nYg6jzotiAc&t=893s>
2. <https://www.nist.gov/pml/owm/metric-si/si-units>
3. <https://www.nist.gov/pml/special-publication-811>
4. <https://www.physics.lbl.gov/rememberinggeoffreychew/>

Chapter 2

Building blocks: Time and space

If we want to describe the motion of a material point, we give the values of its coordinates as a function of time. However, we should keep in mind that for such a mathematical description to have physical meaning, we first have to clarify what is to be understood here by “time”. We have to bear in mind that all our propositions involving time are always propositions about simultaneous events. If, for example, I say that “the train arrives here at 7 o’clock”, that means, more or less, “the pointing of the small hand of my clock to 7 and the arrival of the train are simultaneous events”.

A. Einstein 1905a

2.1 Time

Time is a primitive quantity. We understand the notion of time intuitively, even if it’s difficult to explain (that’s why it’s taken as primitive). In 1905, with the theory of relativity, part of our everyday intuition about this notion was seriously shaken. For many years afterwards our old intuition could still be used in practice and in applications. But the new, correct intuition is becoming more and more important in everyday life and technologies. For example, GPS navigation – which we use everyday from leisure activities such as hiking or sightseeing to more critical ones such as aeroplane landing – critically depends on the correct notion and intuition of time.

Let’s see how our traditional intuition goes astray with a concrete experiment. Here’s Alice, Bob, and Charlie. They have extremely precise

clocks built in exactly the same way. They stay very close to one another and synchronize their clocks. Still keeping close, they go around, maybe on an aeroplane or space ship, and all the time they check their clocks. They notice that their clocks stay perfectly synchronized all the time, no matter where they go and what they do.

At some point they separate, each one going around independently. One of them might stay in place, another might take a helicopter, and another might go for a trip on Mars and back.

Alice and Bob at some point meet again, and compare their clocks. They see that their clocks aren't synchronized anymore; the difference could be as small as microseconds, or as large as years. In fact, if this time discrepancy is large, they would notice that they themselves have aged differently; so time discrepancy doesn't affect the clocks only. Let's say for concreteness that Alice's clock is ahead of Bob's, or equivalently that Bob's is behind Alice's. Note the following aspects:

First, neither Alice or Bob can say "my clock was wrong": neither has noticed anything strange about the "passage of time".

Second, they might wonder what's the time on Charlie's clock. But Charlie is at some distance away. They could decide to contact Charlie via radio, say, and ask "what shows your clock right *now*?". But they would notice that there's a delay, even if extremely small, in the radio transmission; so it's unclear to what time would Charlie's answer apply. If we say "let's account for the radio-signal speed", we see that there's a logical problem: speed is distance divided by time, and here we have a problem in exactly determining what's the "correct" time! So we would be reasoning in circles. Besides, even neglecting these difficulties, Charlie's answer could reveal a time that completely different from Alice's and from Bob's – it could be years ahead or behind both of theirs!

Third, if they now stay together, they will see that their clocks remain exactly synchronized, besides the discrepancy they noticed when they met. This discrepancy doesn't increase or decrease. They may even retrace together Alice's and Bob's previous trips; their clocks still remain synchronized.

The experience just described will occur again any time two or more of them meet. There could be a hundred observers like Alice, Bob, Charlie, initially at the same place and synchronized. Whenever two or more of them meet after having been separated, they will notice discrepancies in their clocks. But their clocks will have exactly the same time lapses as long as they stay together. See the illustration in fig. 2.1

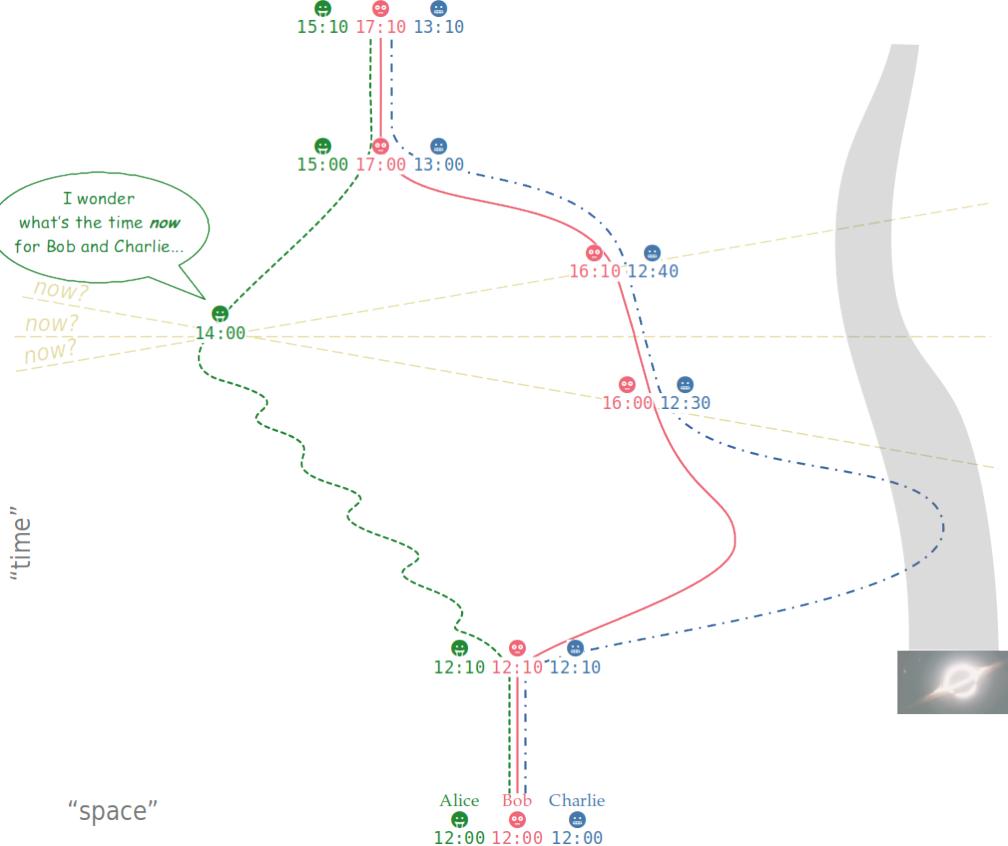


Figure 2.1 Illustration of the experiences of Alice (dashed 😎), Bob (solid 😊), Charlie (dot-dashed 😌) with time. The page represents a two-dimensional spacetime, and is followed from bottom to top.

Bottom: Alice, Bob, Charlie stay close and observe their clocks are perfectly synchronized from 12:00 to 12:10, then they separate.

Right: Charlie visits a region near a strong mass-energy source. Upon meeting again with Bob, the two notice their clocks differ: 16:00 for Bob, 12:30 for Charlie. Yet this clock difference stays the same while they travel together for 10 min.

Left: Alice wanders around travelling at high speed with respect to the fixed stars. At some point she wonders what's the time "now" for Bob and Charlie. Obviously this question doesn't make sense, because (1) when Bob and Charlie are together their clocks differ – impossible to say what's "the" time at their position; (2) not clear which instant in Bob & Charlie's trajectory should be considered as "now" (yellow dashed lines).

Top: When all three meet again, their clocks have completely different readings; and they themselves have aged differently. But their clocks run again at the same rate as long as they stay close.

Consider for a moment an imaginary world in which these experiments had given a different kind of result. According to Newtonian mechanics, whenever two or more initially synchronized observers like Alice, Bob, Charlie had met, their clocks would have always shown exactly the same time. If one year, 23 days, 8 hours, 9 minutes, and 3.045 399 283 240 992 663 02 seconds have passed for you since you last met Alice, you'd see that exactly the same amount of time has passed for her when you two meet again. If you think about it, in this case it would have been somewhat natural to think "right now, the clocks of far-away Alice, Bob, Charlie must show the same time as mine" (even though you have no real experimental way of confirming that).

But that's an imaginary world. In our world is the more complicated situation described initially that holds. Only one conclusion can be drawn from these experimental results: **Time is not some sort of universal quantity. It is, so to speak, "local" to a person or clock, or to a group of persons or clocks that stick together.** This also means that *it doesn't make sense to ask questions like "what can be the time for far-away Charlie, right now?"*

The time measured by a specific observer is called the **proper time** of that observer. Luckily we know more about how the proper times of separated observers can differ when they meet again. It turns out – according to our current understanding – that the time differences depend, roughly speaking, on how fast the observers are moving with respect to one another and to matter around the universe, and on how much energy is contained in the regions they travel. The general theory of relativity gives us the equations determining any such proper-time differences.

The situation depicted in the experiments above is real. It can be measured, for example, comparing initially synchronized clocks that have been put in aeroplanes flying in different directions. Most importantly, it affects everyday relevant technologies such as the Global Positioning System. Formulae from general relativity appear in your phone's GPS software; see for instance § 20.3.3.3 of the Interface Control Document IS-GPS-200 at <https://www.gps.gov/technical/icwg/>. It must also be taken into account in the establishment and synchronization of time in our everyday equipments:

International Atomic Time (TAI) is based on more than 250 atomic clocks distributed worldwide that provide its stability, whereas a small number of primary frequency standards provide its accuracy. Universal Coordinated Time, which is the basis of all legal time scales, is derived from TAI. To

"In 1976, the International Astronomical Union introduced relativistic concepts of time and the transformations between various time scales and reference systems. [...] Now [...] it is necessary to base all astrometry, reference systems, ephemerides, and observational reduction procedures on consistent relativistic grounds. This means that relativity must be accepted in its entirety, and that concepts, as well as practical problems, must be approached from a relativistic point of view."

Kovalevsky & Seidelmann 2004

allow the construction of TAI and the general dissemination of time, clocks separated by thousands of kilometres must be compared and synchronized. [...] The achieved performances of atomic clocks and time transfer techniques imply that the definition of time scales and the clock comparison procedures must be considered within the framework of general relativity.

(Petit & Wolf 2005)

In most everyday situations for us, who live on or nearby Earth and move at speeds much smaller than c with respect to one another, the discrepancies between our proper times are so small that cannot be measured with ordinary clocks or with our internal clocks. Consider a person walking 10 m away from you and then immediately walking back to you, at 1 m/s. The time elapsed for you will be 20 s, but for that person will be 19.999 999 999 999 999 889 s, a difference of 10^{-16} s, which is the error of an atomic clock. If human beings still exist in some decades or centuries, with space travel they will probably have to deal more and more with proper-time discrepancies also in everyday life.

For the most part of the rest of these notes, we won't need to deal with differences in proper time. But I recommend that you keep present how time really works, and that these small time discrepancies exist and occur all the time along your *worldline*.

Time has physical dimension of time and we shall for the most part measure it using the unit *second*, symbol 's'.

2.2 Space

Together with the notion of time, also the notion of space loses some of its traditional intuition. Several observers in motion with respect to one another will generally disagree on the dimensions of an approximately rigid object in their vicinity. For objects that are far away from an observer, the very notion of "distance" becomes tricky has different and non-equivalent definitions; one must be very careful on which definition is being used.

We shall not delve further into these peculiarities of time and space. Keep simply in mind that phenomena happen in **spacetime**, and that there's no way to attribute a universal time, nor a universal position in space, to a physical event. There is one absolute: whoever locally measures the speed of light, will find the value

$$c := 299\,792\,458 \text{ m/s} \quad (2.1)$$

"The plot for Cesium [...] characterizes the best orbiting clocks in the GPS system. What this means is that after initializing a Cesium clock, and leaving it alone for a day, it should be correct to within [...] 4 nanoseconds. Relativistic effects are huge compared to this."

Ashby 2003

This value is exact by definition, and serves as a way to define a local notion of space and distance.

Space has physical dimension of length and we shall for the most part measure it using the unit *metre*, symbol ‘m’.

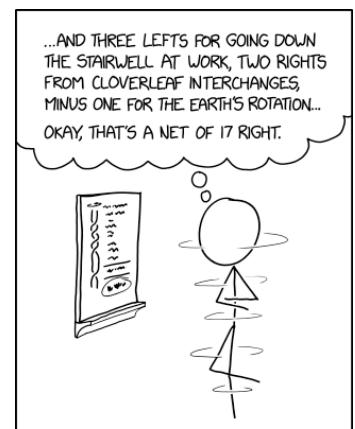
2.3 Coordinate systems and events

It is necessary to have a way for distinguishing physical events and phenomena and locating them in spacetime. This is achieved through a **coordinate system**.

A coordinate system assigns four numerical labels to every point in spacetime. Often these labels have some kind of physical meaning – such as the proper time elapsed for a specific clock, or the distance from some event as measured by a specific observer – but they don’t need to.

A coordinate system also solves the problems coming from proper-time and space discrepancies among different observers. We can assign to every physical event a *coordinate time* and a *coordinate spatial position*, which are the same for all observers, because decided by agreement. Coordinate time doesn’t have a strict physical meaning, and will generally be different from the proper times registered by different observers. It can nevertheless be used for “doing physics”, and it is the time we shall most often use in our equations. A coordinate time commonly used for Earth-physics purposes is [Universal Coordinated Time \(UTC\)](#)¹. The clock on your phone, and on devices that get synchronized via internet, shows UTC, not your proper time. An observer on Earth at 0 m over sea level, and not moving, measures a proper time exactly equal to UTC (besides small variation coming from the movements of Solar System bodies). Observers at other altitudes or moving with respect to Earth’s surface notice that their proper times are slightly different from UTC.

Up to now we have often used the word ‘event’, informally taking its meaning for granted. Let’s be more precise: we call **event** or **spacetime point** an extremely small region of space – a point – that only lasts for a very small lapse of time – an instant. The name ‘event’ is used because typically we approximately identify such a point and instant by means of a physical phenomenon of limited spatial extension and short duration, such as the collision of two subatomic particles. The sudden burst of a very small soap bubble can be considered as an event in some circumstances; but something like “a tennis ball” cannot be considered as an event, mainly because a tennis ball exists for quite a long time, not just for a short instant.



From a four-dimensional spacetime point of view, a tennis ball could be characterized as a line: a **worldline**.

We shall often denote the four coordinates of a coordinate system by the letters

$$(t, x, y, z)$$

where t is a coordinate time, usually UTC, and (x, y, z) determine a spatial position. The triplet of spatial coordinates is often denoted by the vector \mathbf{r} :

$$\mathbf{r} := (x, y, z) \quad \text{or} \quad \mathbf{r} := [x, y, z] \quad \text{or} \quad \mathbf{r} := \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

use round brackets ‘()’ or square brackets ‘[]’, and horizontal or vertical notation as you prefer.

It is always important to specify how the coordinate system you’re using is defined. The definition of the spatial coordinates (x, y, z) is typically different from problem to problem. We shall typically use coordinates that form $\frac{\pi}{2}$ rad $\equiv 90^\circ$ angles with one another; but their directions and their origin – that is, where they have value $x = y = z = 0$ m – always depend on the problem, so make sure you always specify them.

Whenever we speak of a “region of space” or of a “surface in space”, we mean a 3D or 2D region at some specific coordinate time t .

Some physical phenomena happen approximately along a line, in one dimension; or on a surface, in two dimensions. In these cases we can omit two or one of the spatial coordinates, assuming they have some constant, unimportant values; and we can simply write, for instance, (t, x) or (t, x, y) as our coordinates.

Exercise 2.1

Consider a clock at rest on the Earth’s surface, at a distance r_e from the Earth’s centre; and a clock on a satellite, for instance a GPS satellite or the [International Space Station](#)², right above the first clock, at a distance r_s from the Earth’s centre. An observer by the clock on Earth measuring a time lapse Δt_e will see the clock on the satellite has having run for a time lapse Δt_s , and vice versa (note that this “vice versa” only holds in this specific situation!). The relation

between two time lapses is approximately given by

$$\frac{\Delta t_s}{\Delta t_e} = \frac{\sqrt{1 - 2\frac{G}{c^2} \frac{M}{r_e}}}{\sqrt{1 - 2\frac{G}{c^2} \frac{M}{r_s}}}$$

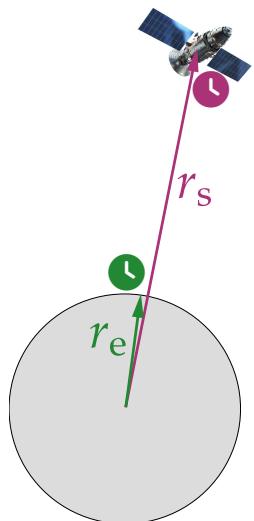
where $G \approx 6.7 \times 10^{-11} \text{ m}^3/(\text{kg s}^2)$, $c = 3.0 \times 10^8 \text{ m/s}$, and the Earth's mass $M = 6.0 \times 10^{24} \text{ kg}$.

1. Take the case of a GPS satellite, with $r_e = 6.4 \times 10^6 \text{ m}$ and $r_s = 2.6 \times 10^7 \text{ m}$ ([NASA data³](#)). If you, on the ground, measure a time lapse of $\Delta t_e = 10 \text{ years}$, what's the difference, in seconds, with the time lapse Δt_s you see on the satellite?
2. If the time lapses are large compared with the time needed to go from ground to orbit or vice versa, then $\Delta t_s/\Delta t_e$ is also the ratio between the real *ageing* of a person who's been in orbit and one who's been on the ground, when they meet again.

Now consider the case with a black hole instead of Earth. The formula above still apply approximately.

In the movie *Interstellar⁴*, two astronauts go on Miller's planet, at a distance r_e from the black hole Gargantua, and stay there for 3 hours, leaving one astronaut in orbit at a distance $r_s \approx \infty$ (the distance is large enough that it can be approximated as infinity). When they meet again, the latter astronaut has aged 23 years.

Given that Gargantua's mass is $M = 2.0 \times 10^{38} \text{ kg}$, calculate the distance r_e of Miller's planet from the black hole.



2.4 Velocity and acceleration

In some situations the spatial coordinates $\mathbf{r} = (x, y, z)$ may turn out to be functions of the time coordinate t ; the typical example is when we describe how the spatial position of a small body changes with coordinate time. We can write this functional dependence in different ways, for instance

$$\mathbf{r}(t) \quad \text{or} \quad [x(t), y(t), z(t)].$$

So \mathbf{r} is a vector function of time, which simply means that we have a collection of three functions of time.

If we take the derivative of each coordinate with respect to the time t , we obtain the **coordinate velocity**

$$\mathbf{v}(t) := \frac{d}{dt} \mathbf{r}(t) = \left[\frac{d}{dt} x(t), \frac{d}{dt} y(t), \frac{d}{dt} z(t) \right]$$

which is also a vector.

Dot-notation for time derivative

The derivative of some quantity with respect to coordinate time is often denoted by a **dot** over the quantity. So we can also write

$$\mathbf{v}(t) = \dot{\mathbf{r}}(t) = [\dot{x}(t), \dot{y}(t), \dot{z}(t)]$$

The coordinate velocity is usually different from the *physical velocity*, which an observer would measure using proper time and space, for instance using bouncing light rays. In many everyday situations the difference between coordinate and physical velocity is so small that it can be neglected. But in situations involving subatomic particles at high speed, for example, one must take into account that the two velocities are different.

Taking the time derivative once more we obtain the coordinate acceleration, also a vector:

$$\mathbf{a}(t) := \frac{d}{dt} \mathbf{v}(t) = \frac{d^2}{dt^2} \mathbf{r}(t) = \left[\frac{d^2}{dt^2} x(t), \frac{d^2}{dt^2} y(t), \frac{d^2}{dt^2} z(t) \right]$$

which is also a vector.

Acceleration in relativity theory

In relativity theory, acceleration acquires a special physical significance, because it includes the effect of gravity, and its calculation does not involve just a time derivative. For instance, let's say that you are standing still on the ground, and let's use a coordinate system where x points in front of view, y to your left, and z points upwards. Then your coordinate velocity is $\mathbf{v} = (0, 0, 0)$ m/s also according to relativity theory. But the spatial part of your acceleration is approximately $(0, 0, 9.8)$ m/s², not zero!

The definitions and values of acceleration according to relativity theory and according to Newtonian mechanics are therefore quite different even in everyday situations. In these notes we'll mean the Newtonian definition of acceleration, unless stated otherwise.

 **Exercise 2.2**

1. Here are the three components of a time-dependent velocity vector; the variable t is the time, and therefore has dimension time. Introduce units 's' and 'm' appropriately in such a way that the expression is dimensionally correct:

$$\mathbf{v}(t) = [4, \cos(3t), -\exp(8/t)]$$

2. The position vector of a satellite is given below. Calculate the satellite's velocity and acceleration vectors:

$$\mathbf{r}(t) = \begin{bmatrix} 2.0 \times 10^7 \cos\left(\frac{t}{13751}\right) \\ 2.2 \times 10^7 \sin\left(\frac{t}{13751}\right) \\ 0 \end{bmatrix} \text{ m}$$

3. What is the satellite's velocity at $t = 6875$ s? What is the magnitude of the velocity (that is, the speed)?

URLs for chapter 2

1. [https://www.nist.gov/pml/time-and-frequency-division/time-realization/utc
nist-time-scale-0/](https://www.nist.gov/pml/time-and-frequency-division/time-realization/utc-nist-time-scale-0/)
2. <https://www.nasa.gov/international-space-station/>
3. [https://www.nasa.gov/directorates/somd/space-communications-navigation-p
rogram/gps/](https://www.nasa.gov/directorates/somd/space-communications-navigation-program/gps/)
4. <https://www.imdb.com/title/tt0816692/>

Chapter 3

Building blocks: “Stuff”

if the skill of the mathematician has enabled the experimentalist to see that the quantities which he has measured are connected by necessary relations, the discoveries of physics have revealed to the mathematician new forms of quantities which he could never have imagined for himself.

J. Clerk Maxwell 1870

3.1 Seven primitive quantities with three basic properties

Besides time and space, our physics formalism includes around seven more quantities that we take as primitive:

Seven primitive quantities

matter
electric charge
magnetic flux
energy
momentum
angular momentum
entropy

Technically they are called *fields*, for reasons we shall shortly see.

Recall that primitive quantities cannot be defined: we can only try to understand them intuitively, for example through their properties. Some of the seven primitive quantities are easier to grasp intuitively than others.

But all seven primitive quantities have three basic properties in common. First of all, of each we can ask, or rather, measure:

The two basic measurements that can be made on the seven quantities

1. How much of this quantity is in a particular three-dimensional region of space at a particular time instant?
2. How much of this quantity flows through a particular two-dimensional surface during a particular time lapse?

We can ask these questions of any region of space and any time lapse, and the surface in the second question can be moving and deforming. The results of the two measurements above are numbers, which in general can be positive or negative, for scalar quantities; or vectors for vector quantities.

Often we consider the second question, about the flow through a surface, in the case of a very short lapse of time, and divide the total flow by that time lapse. So we have an alternative form of the second measurement:

Flux of a substance through a surface

- 2b. How much of this quantity is flowing through a particular two-dimensional surface per unit time, at a particular time instant?

This is called the **flux** of the quantity through that surface.

The third property common to the seven quantities regards the two measurements above.

Extensivity of the seven quantities

If we consider two or more separate volumes, the amount of quantity in the total volume is equal to the sum of the amounts in the separate volumes.

Analogously for the flux through separate surfaces.

We say that each of the seven quantities is an **extensive** quantity

The basic measurements above can't in general be made, and don't even make sense, for some other quantities. For instance, we cannot ask "what's the total amount of *temperature* in this region?", or "how much velocity is flowing through this surface?".

Thanks to the three properties above, each of the seven quantities can be intuitively visualized as some kind of “stuff” that can be present at each spacetime point, fills regions of space, and flows through surfaces. This visualization is useful, but also comes with some warnings which we’ll discuss later.

What’s remarkable about matter, electric charge, magnetic flux, energy, momentum, angular momentum, and entropy, is that *they are common to all our main physical theories*, approximate or not: from Newtonian mechanics to general relativity and quantum theory; from subatomic scales to cosmological scales. The physical meaning and mathematical characterization of these quantities can be slightly different depending on the physical theory and scale. For example, in quantum theory they are mathematically represented by so-called operators rather than functions; and at molecular scales entropy has a meaning connected with probability theory. Yet, these seven quantities are really universal to our present way of doing physics and of describing physical phenomena around and within us.

Let us make a first acquaintance with these seven quantities. The discussion that follows is meant as an introduction; we shall repeat and say more about each quantity in later chapters.

Keep in mind that it is very difficult, if not impossible, to answer questions like “what *really* is...?” or “why *really* does...?”.

3.2 Matter

Matter is a *scalar* quantity, with SI dimension [amount of substance²](#), and measured in units of [moles \(mol\)³](#). Its flux is therefore measured in units of *moles per second* (mol/s). In statistical mechanics and particle physics, matter is often simply counted and so measured in dimensionless units, rather than moles.

Matter is probably the easiest quantity to grasp intuitively; what we ordinarily call “stuff” is matter. It is usually classified into several kinds; the classification depends on the physical phenomena and physical theory one works with, and is related to whether the kinds of matter can be considered separately conserved, as we’ll discuss in § 6.

In everyday phenomena not involving [radioactivity⁴](#) or [nuclear energy⁵](#), the different kinds of matter approximately correspond to the non-radioactive [chemical elements⁶](#): [hydrogen⁷](#), [helium⁸](#), [lithium⁹](#), and so



Richard Feynman explains the difficulty of “why” questions in a funny and insightful video¹.

on. Note that some common everyday devices, such as smoke detectors, do involve radioactivity.

In phenomena involving radioactivity or nuclear energy, the different kinds of matter correspond approximately to *baryons*¹⁰, such as protons and neutrons; and *leptons*¹¹, such as electrons. In particle physics, even more subtle classifications of matter are made, into kinds that seem to be conserved, such as electronic-leptons, muonic-leptons, and others. This kind of research is still open, but it seems that the total amount of baryonic and leptonic matter, independently of the kinds into which it can be classified, is always conserved. Note that we're using the term 'matter' in a sense that includes anti-matter, such as positrons.

The total amount of matter in a region *can be negative*. "Negative matter" is what's usually called *anti-matter*¹². Anti-matter appears in small amounts in everyday life, for example in connection with common radioactivity processes. It is also created and used in medicine, in *positron-emission tomography (PET)*¹³ scans.

In these notes we shall usually not consider distinctions between different kinds of matter, making some exceptions for discussions about chemical reactions and nuclear phenomena.

Matter: notation

The amount of matter in a region is usually denoted with N , and flux of matter with J . In chemistry we usually specify what kind of matter we are speaking about, writing for instance $N_{\text{Ca}} = 5.3 \text{ mol}$, to indicate an amount of 5.3 mol of *calcium*¹⁴ atoms.

Matter is different from mass or energy

It is important to clearly distinguish matter from *mass* or *energy*. Mass can be considered a property of matter, but the two are different. In nuclear reactions, for instance, the mass of some amount of matter may change, even if the amount of matter stays the same.

As far as we know, the total amount of energy associated with an amount of matter is always positive, whether the amount of matter is positive or negative (antimatter). This is the reason why antimatter "falls" just like positive matter, a fact that has been experimentally confirmed: see Anderson et al. 2023.

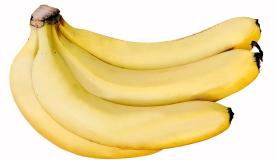
Exercise 3.1

According to statements on symmetrymagazine.org¹⁵ and quantum-diaries.org¹⁶,

The average banana (rich in potassium) produces a positron roughly once every 75 minutes.

Unfortunately the original site where this statement was discussed, and the corresponding calculation made, seems not to exist anymore.

1. Do a little research and find out whether this statement is true.
2. From your research, approximately quantify the flux of positrons around an ordinary banana, expressing it in particles/s.



How many positrons do bananas produce?

3.3 Electric charge and magnetic flux

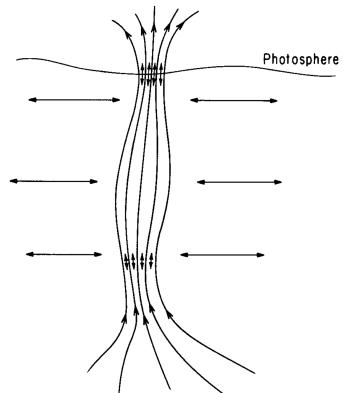
Electric charge is a *scalar* quantity, with SI dimension [electric charge](#)¹⁷ (equivalent to electric current×time), and measured in units of [coulombs](#) ([C](#))¹⁸. Flux of electric charge is called *electric current*, and measured in units of [amperes](#) ([A = C/s](#))¹⁹.

Electric charge is a quantity that is easily grasped in our everyday experience, and doesn't need much comments.

Magnetic flux is a *vector* quantity, with SI dimension [magnetic flux](#)²⁰, and measured in units of [webers](#) ([Wb](#))²¹.

The electromagnetic field is most commonly represented by vectors associated to each point in space. But it can also be interpreted and visualized as a collection of moving, oriented tubes or lines, either closed or extending indefinitely; somewhat analogously to how we visualize matter and charge, as moving blobs or points, but with one more dimension. This interpretation goes back to Faraday (1846), Maxwell (1855), and later Dirac (1955) among others, and today is conveniently used in some fields such as [solar physics](#)²², for example to study [sunspots](#)²³ (see Ryutova 2018). From this point of view, the *voltage* turns out to be the flux of the magnetic flux; it is indeed measured in *volts* (V), equivalent to webers/second.

Electromagnetism and this particular visualization of it are very fascinating topics, but we shall not discuss them in these notes.



"sketch of the magnetic lines of force in a magnetic filament extending up through the photosphere." Parker 1974a

3.4 Energy

Energy is a *scalar* quantity, with SI dimension energy, and measured in units of *joules* (J); its flux is measured in units of *watts* (W = J/s).

Equivalently we can speak of mass, with SI dimension mass, and measured in units of *kilograms* (kg)²⁴; its flux is measured in *kilograms per second* (kg/s).

The notion of energy is extremely important today, and central in many world-wide important discussions and worries – think of the “energy crisis”, “renewable energy”, and so on. It is somewhat funny that despite this importance it’s actually difficult to answer ‘what *is* energy, really?’. Often we speak about energy as something that “flows”, is “transported”, “converted”, “stored”, and similar visualizations. This intuition will be enough in these notes. The notion of *mass* is also very intuitive in our everyday life; we associate it with the difficulty in setting objects into motion, or with the weight of objects.

From Relativity Theory and experimentally we know that *energy and mass are the same quantity*, and in these notes we shall emphasize this experimental fact.



Energy and mass are the same

Let’s see some examples of why it is impossible to make a real distinction between energy and mass. The following examples have been simplified in some of their aspects, but their main point is valid.

Heated gas Imagine we have a box with a given amount of gas, say 1 mol of oxygen molecules. Using an extremely precise weighing scale, we observe that the mass of the gas is, say, exactly

$$0.031\,999\,540\,000\,000\,000 \text{ kg} .$$

Now we heat the gas, providing 60 J, while making sure that not a single molecule of oxygen gets in or out of the box. The temperature of the gas increases by around 3 K. We observe that the weight measured by the scale increases while we heat the gas, reaching the new value

$$0.031\,999\,540\,000\,000\,668 \text{ kg} .$$

Clearly the mass has increased, but no molecules were added! The additional mass is the 60 J that we provided to the gas. Energy has weight, energy is mass.

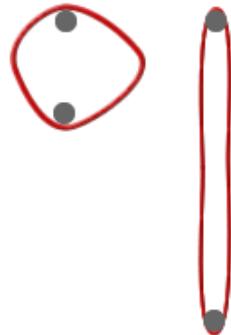
Stretched or moving rubber band Take a common rubber band, and imagine again that we have an extremely precise weighing scale. The rubber band, unstretched, has a mass of exactly

0.000 500 000 000 000 000 000 kg .

Now we stretch the band a little. According to Newtonian mechanics it acquires ‘elastic energy’, which increases its initial internal energy by, say, 0.3 J. Now we weigh the rubber band again, while stretched, and observe a mass of approximately

0.000 500 000 000 000 003 338 kg .

The extremely small difference of around 3×10^{-18} kg from the initial mass is exactly the elastic energy that we provided by stretching. Energy has weight; energy is mass.



When we stretch a rubber band, its mass increases slightly – even if the amount of rubber remains exactly the same.

Now set the unstretched band in motion, so that according to Newtonian mechanics it acquires a kinetic energy of, say, 0.2 J. If we could weigh the band while in motion (but without moving the weighing scale), we would measure a mass of approximately

0.000 500 000 000 000 002 225 kg .

The small difference from the initial mass is the new kinetic energy of the band.

This case is actually connected with the example of the gas above. If we observed the gas at a molecular level, we would interpret the energy of 60 J provided to it as additional kinetic energy of its molecules. The increase in weight was exactly this additional kinetic energy.

Fission and atomic bombs The atomic bomb²⁶ is a dark example of the fact that mass is energy. In the case of nuclear fission, if we weigh the amount of nuclear material, say within a box, before and after fission, we observe that its mass has decreased. But we also observe that a great amount of energy has been released out of the box.

Electric heater As a final example consider a 1000 W electric heater, which is radiating 1000 J in one second. The heater is also losing around 0.000 000 000 01 kg of mass every second owing to this heat radiation – although it's also acquiring the same amount of mass as electromagnetic energy.



Hydrogen Bomb Test, 1954 (National Museum of Nuclear Science & History²⁵)

The equivalence between energy and mass is given by the famous formula $E = mc^2$, where c is the speed of light, eq. (2.1). In their respective units this gives

$$1 \text{ kg} = 89\,875\,517\,873\,681\,764 \text{ J} \quad (\text{exactly})$$

$$1 \text{ J} \approx 0.000\,000\,000\,000\,000\,011\,126\,5 \text{ kg}$$

To grasp these numbers, consider that the mass of the rubber band in the example above, 0.5 g, is comparable to the energy released by the [atomic bomb over Hiroshima](#)²⁷.

"we are led to the more general conclusion: The mass of a body is a measure of its energy content; if the energy changes by L , the mass changes in the same sense by $L/9 \cdot 10^{20}$, if the energy is measured in ergs and the mass in grams." Einstein 1905b

The practical used of the words 'mass' and 'energy'

From the examples above it becomes clear that energy and mass are two names for the same thing. But it also becomes clear that in our daily experience we deal with energy-mass in two different ways:

On the one hand, we deal with huge (atom-bomb-like) amounts of energy packed in very small volumes: the huge amount of energy that goes together with objects such as pens, keys, bicycles, cars, houses, and so on. We move, push, pull these huge energy amounts from one place to another, and even put them in our pockets. These energy amounts change a little all the time; see the examples above. But the change is so small as to be often undetectable with ordinary scales, and negligible for practical purposes. We call this energy 'mass' and measure it with a unit – kg – that doesn't lead to ridiculously large numbers. And we also agree to neglect the imprecision and fluctuation in its measurement, say any imprecision under 0.000 001 %.

On the other hand, we also deal with the small energy changes and exchanges in all these objects. These exchanges that are very important for our daily life: they keep us warm, make our cells work, make our laptops work. In dealing with these energy exchanges, we don't care about the huge energy reservoirs they come from. So we agree to measure them with a unit – J – that doesn't lead to ridiculously small numbers. And we also agree not to be precise about the total amount in the reservoir from which these energy bits come from.

As an analogy, consider when we speak about the amount of people in different countries. We can say that in Norway there are 5 millions, and in India 1500 millions, so in India there are 300 times more people. By this we don't mean that in Norway there are *exactly* 5 000 000 people and

that India has *exactly* 300.000 000 times more people. These numbers are changing slightly all the time; but we don't care about differences of 10 or even 10 000 people. At the same time, think of when you have three dear friends or relatives visiting you from abroad: the amount of 3 people is now for you very important; and you don't care about how much this amount is compared to the total population of your country.

The distinction above is of course not clear-cut. In dealing with some physical phenomena, for example with few molecules or in particle physics, the distinction become too blurry and not useful anymore. And indeed in these situations one often uses the terms 'mass' and 'energy' interchangeably, as well as a common unit for both (for instance *electronvolts*²⁸).

In these notes we shall often use the expressions 'energy-mass' and 'mass-energy' to remember that these two words denote the same thing.

Different 'forms' of energy

We often speak of different *forms* of energy-mass. The most important for us will be **internal energy**, **kinetic energy**, **gravitational potential energy**, to be discussed later; another important one is *electromagnetic energy*.

In ch. 5 we shall see that the differences among these energy forms come from the way they are calculated from other quantities, such as matter or magnetic flux and electric charge. For example, if we know that in a volume there's an amount of a particular kind of matter, then we know that there must also be an amount of energy given by a particular formula. And if that matter is moving, then we have to add to the total an extra amount of energy given by another formula. And if in that volume there's a gravitational field, then another extra amount must be added, given by yet another formula. Similarly if we know that an electromagnetic field is in that volume.

We also speak of different forms of flux of energy. The most important for us will be **heat** and **work**, and possibly **convection**. The difference is again in how these fluxes are calculated depending on whether there are also fluxes of matter.

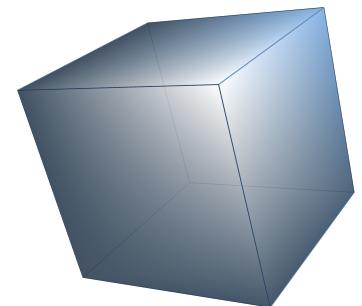
Amounts and forms of energy are coordinate- and observer-dependent

An aspect of energy that must always be kept in mind is that **the amount of energy depends on the coordinate system we're using**. If someone points at a specific region of space at a particular instant, and asks "how

much energy is there?", we *cannot* give an answer until a coordinate system is specified. Once the coordinate system has been chosen, then a precise and unambiguous answer can be given. The same is true for the flow of energy through a surface. This also means that observers using different coordinates will usually assign different amounts of energy to the same regions of spacetime.

This is an important difference between energy on one side, and matter and electric charge on the other side. *For matter and electric charge, the questions above can be answered unambiguously independently of any coordinate system.* But not so for energy. The reason of this quirky difference is ultimately connected with the fact that time and space are also observer-dependent.

This coordinate-dependence is not a problem: we must always specify our coordinate system anyway, in order to agree on the time and position of physical events. But it can cause problems when we calculate *changes* in energy. If we first calculate or measure some amount of energy in a coordinate system, then we calculate or measure another amount at some later time in a *different* coordinate system, the difference between the two amounts has no meaning whatsoever. In fact it could happen that there's no energy change at all in one coordinate system or in the other: any change we found was just an artefact of mixing up coordinates.



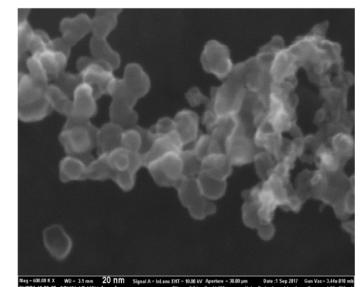
"How much energy is there in this volume at this instant?" This question cannot be answered until we have specified which coordinate system we're using.

Amounts of energy-mass are coordinate-dependent

Never change coordinate system in the middle of energy calculations!

Also the distinction among different forms of energy is coordinate-dependent. For instance, in one coordinate system we can say that a given volume contains only internal energy; but in another coordinate system that same volume can be said to contain internal and kinetic energy.

The distinction among different forms of energy also depends on the observation scale and on the theory used. For instance, we can observe and model the gas in a container on a scale of metres, seeing it as a uniform flux; in this case we say that there's internal energy in the container. But if we observe and model the gas as a collection of molecules, on a microscopic scale, then we say that there's internal energy (of the single molecules) and kinetic energy in the container (the total energy being the same, unless we have changed coordinate system).



Hydrocarbon fuel particles²⁹. The small blobs have size of around 2×10^{-8} m.

The same is true of energy flux: what we call ‘heat’ on one observation scale is ‘work’ on a finer scale.

Energy-mass: notation

The amount of energy in a region is usually denoted with E , or with M if we describe it as mass. Internal energy is denoted U , kinetic energy E_k , potential energy E_p .

The total flux of energy will be denoted by Φ ; the flux in the form of heat, by Q ; and in the form of work, by W .

Exercise 3.2

In an hour, 14 people exit through a door. Taking the average human weight to be 62 kg (Walpole et al. 2012), what’s the average *energy* flux, in J/s, through that door?

3.5 Momentum

Momentum, also called *linear momentum* or *translational momentum* to distinguish it from angular momentum, is a *vector* quantity. Its SI dimension and units can be written in several equivalent ways; we shall keep in mind especially these three:

$$\begin{array}{ccc} \text{force} \times \text{time} & \equiv & \text{mass} \times \text{length/time} \\ \text{N} \cdot \text{s} & & \text{kg} \cdot \text{m/s} \end{array} \quad \equiv \quad \begin{array}{c} \text{energy} \times \text{time/length} \\ \text{J} \cdot \text{s/m} \end{array}$$

Since it is a vector quantity, it is usually expressed with three numbers, typically the x -, y -, and z -components. In simplified problems where only one or two dimensions are relevant, only the relevant components are reported.

Momentum is a subtle quantity, even subtler than energy. Textbooks that focus on Newtonian mechanics *define* it as the product of the mass and the velocity of a body, usually written “ $\mathbf{p} = m\mathbf{v}$ ”. This relation, however, is only valid in special circumstances, and cannot be used in many everyday technological applications, especially when electromagnetism is involved. And that relation is actually only an approximation even in the circumstances where it’s used.

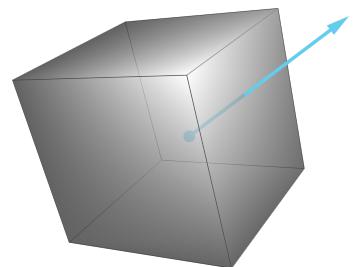
It is therefore convenient to separate our idea of momentum from the idea of “objects” moving, keeping in mind that the latter idea is just a particular case of momentum. Yet, *momentum is indeed associated with translational motion* of matter and of electromagnetic fields. *Translational motion* is the kind of motion that leads to a *new position* in space. For instance, when you walk from one place to a different one, you have performed translational motion (note that translational motion doesn’t need to be in a straight line).

Just as energy is often mentally visualized as a sort of fluid (although this visualization comes with many warnings), also momentum can actually be visualized as a sort fluid; but you must imagine it as a “fluid of vectors”. Given a particular volume at a particular instant in time, and given a coordinate system, we can speak of the total amount of momentum within that volume. This amount is represented by a *vector*. You can imagine a continuous collection of vectors filling the volume, possibly with different directions and small magnitudes; the total momentum is the sum of all these vectors. This visualization obviously comes with many warnings, but it can be very useful if we are careful.

And just as energy, the total amount of momentum is *coordinate-dependent*. So we have the same warning here:

⚠️ Amounts of momentum are coordinate-dependent

Don’t change coordinate system in the middle of calculations of momentum!



The amount of momentum within a volume is represented by a (3D) vector

Flux of momentum is what we call **force**. A static force, like the one you exert when you hold a bag, is often mentally visualized as a static vector. In ch. 4 we shall discuss a different visualization, in which force is represented as a sort of flow of vectors.

Bookmark Momentum: notation

The amount of momentum in a region is usually denoted with **P** .
The flux of momentum is also called force and denoted with **F** .

🔗 Momentum and energy flux are the same

According to Relativity Theory, momentum *is* energy flux (such as heat), and energy flux *is* momentum. If we represent momentum with the symbol **P** , and energy flux with the symbol **Q** , the equivalence between

them is given by

$$\mathbf{Q} = \mathbf{P}c^2 \quad (3.1)$$

Compare this formula with $E = mc^2$. From this point of view, you can think of momentum as “energy in motion”. This is consistent with our discussion about mass-energy: since mass is energy, the Newtonian expression “ $m\mathbf{v}$ ” indicates energy in motion, or a flux of energy. On a sunny day, if you close your eyes and feel the Sun’s heat on your face, what you are feeling is actually a flow of momentum. And when you kick a ball, you’re setting a huge bundle of energy in motion, and that’s why the ball has acquired momentum.

3.6 Angular momentum

Angular momentum, also called *moment of momentum* or *rotational momentum*, is a *vector* quantity. Its SI dimension and units can be written in several equivalent ways; we shall keep in mind especially these three:

$$\begin{array}{ccc} \text{force} \times \text{length} \times \text{time} & \equiv & \text{mass} \times \text{length}^2/\text{time} \\ \text{N} \cdot \text{m} \cdot \text{s} & \equiv & \text{kg} \cdot \text{m}^2/\text{s} \end{array} \equiv \text{energy} \times \text{time} \quad \text{J} \cdot \text{s}$$

It is usually expressed with three numbers, typically the x -, y -, and z -components.

Angular momentum is probably an even subtler quantity than momentum. Just as momentum is associated with translational motion, angular momentum is associated with *rotational* motion. Rotational motion is the kind of motion that leads to a *new orientation* in space, rather than to a new position. For instance, if you turn to your left or to your right while standing in place, you have performed rotational motion.

There isn’t a clear-cut distinction between translational and rotational motion: usually they involve each other to some degree. A translational motion can be interpreted as a rotation around a point that is very far away; and a rotation of an extended object can be interpreted as small translational motions of its parts.

This is the reason why in many situations we can calculate angular momentum in terms of momentum. If the momentum in a *small* volume is denoted by the vector $\mathbf{P} = (P_x, P_y, P_z)$, and the position vector by

$\mathbf{r} = (x, y, z)$, then the angular momentum $\mathbf{L} = (L_x, L_y, L_z)$ with respect to the origin of coordinates, in that same volume, is given by the vector product

$$\mathbf{L} = \mathbf{r} \times \mathbf{P}$$

or equivalently
$$\begin{cases} L_x = y P_z - z P_y \\ L_y = z P_x - x P_z \\ L_z = x P_y - y P_x \end{cases} \quad (3.2a)$$

Instead of calling the components " (L_x, L_y, L_z) ", we can also call them " (L_{yz}, L_{zx}, L_{xy}) ", as some books do. The latter names make the formulae above easier to remember:

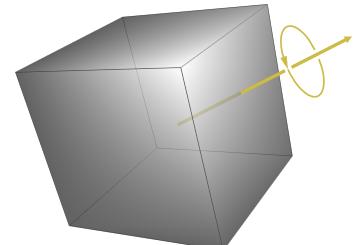
$$\begin{cases} L_{yz} = y P_z - z P_y \\ L_{zx} = z P_x - x P_z \\ L_{xy} = x P_y - y P_x \end{cases} \quad (3.2b)$$

Choose whichever you prefer.

Angular momentum is something that is associated not only with ordinary bodies (matter), but also with electromagnetic fields. Just like momentum, also angular momentum can be visualized as a "fluid of vectors". Given a particular volume at a particular instant in time, and given a coordinate system, we can speak of the total amount of angular momentum within that volume. This amount is represented by a *vector*, and is *coordinate-dependent*:

⚠️ Amounts of angular momentum are coordinate-dependent

Don't change coordinate system in the middle of calculations of angular momentum!



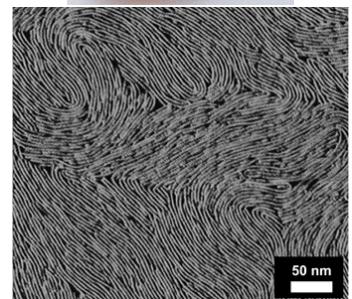
The amount of angular momentum within a volume is represented by a (3D) vector. (Curious about the little circulation symbol around the vector? Then read on.)

The flux of angular momentum is also called the *torque*, and bears a relation to the flux of momentum similar to the formula above. If the force – flux of momentum – is denoted by the vector $\mathbf{F} = (F_x, F_y, F_z)$ and the position vector by $\mathbf{r} = (x, y, z)$, then the flux of angular momentum, or torque, $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)$ with respect to the origin of coordinates is given by

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$$

or equivalently
$$\begin{cases} \tau_x = y F_z - z F_y \\ \tau_y = z F_x - x F_z \\ \tau_z = x F_y - y F_x \end{cases} \quad \text{or} \quad \begin{cases} \tau_{yz} = y F_z - z F_y \\ \tau_{zx} = z F_x - x F_z \\ \tau_{xy} = x F_y - y F_x \end{cases} \quad (3.3)$$

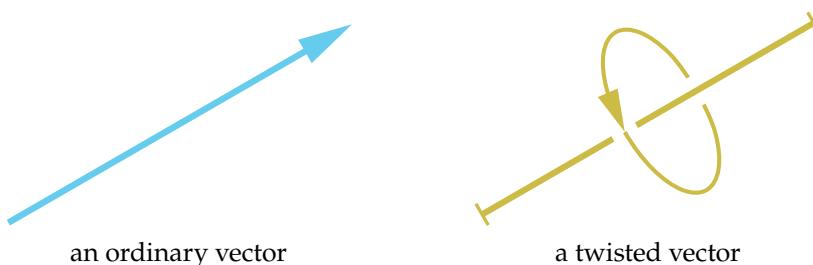
You may wonder: "Do we really need angular momentum? after all it just looks like something constructed from momentum". The answer is yes, we really need it, for two reasons. First, angular momentum obeys an important universal law which is independent from those obeyed by energy and by momentum (Truesdell 1968a tells some of the story of how this was discovered). Second, for some physical phenomena, for example involving liquid polymers³⁰, elementary particles, or electromagnetic radiation, the angular momentum includes an additional part, called *spin* or *intrinsic angular momentum*, that is *not* related to linear momentum. In the present notes we shall not use this more general kind of angular momentum.



Some liquid polymers (**top**: Liquid Diethoxymethane Polysulfide) need to be described with a special kind of angular momentum, owing to their molecular structure (**bottom**).

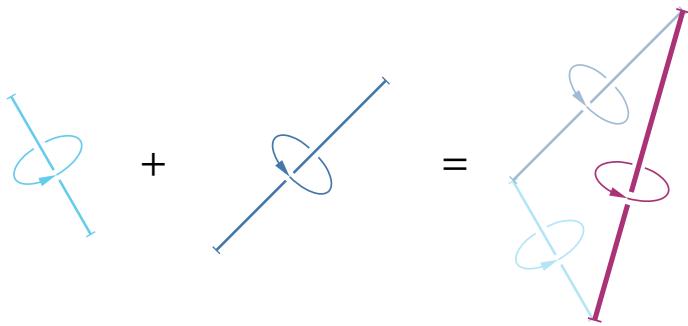
Angular momentum as a twisted vector

In order to represent angular momentum we can use a kind of vectors different from the arrow-like ones (called *polar* vectors) with which you are probably familiar. They are called **twisted vectors**, or also *pseudo-vectors* or *axial* vectors or *outer-oriented* vectors. Twisted vectors represent rotations, and therefore have an orientation, not along them, but *around* them:

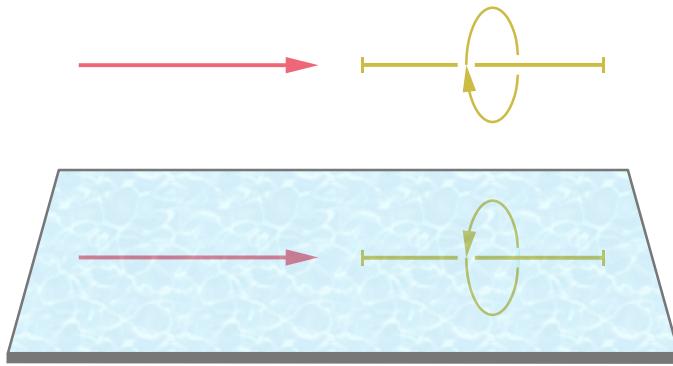


Their length still represents the magnitude of the vector. They make it immediately clear what is the axis of rotation, and what is the sense of rotation.

The sum of twisted vectors is analogous to the sum of ordinary vectors, with the parallelogram rule:



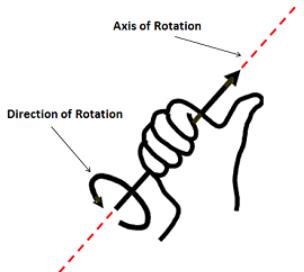
Ordinary vectors and twisted vectors behave very differently if we look at their images through a mirror parallel to their axis: the orientation of ordinary vectors appears unchanged, whereas the orientation of twisted vectors appears *reversed*:



this phenomenon reflects the behaviour of rotations under reflections.

For some mysterious reason many books are afraid of using twisted vectors, and rely on ordinary vectors instead, introducing the “right-hand rule” to determine the sense of rotation from the arrow of the ordinary vector. If you’ve ever asked yourself “why the right hand, and not the left hand?”, the answer is that it’s purely a convention; one could have introduced a left-hand rule instead. Using twisted vectors we don’t need these arbitrary conventions and mnemonics: the sense of rotation is unequivocally indicated by the twisted vector.

Use whichever vector representation you prefer. In these notes we shall use a hybrid notation, as in the side picture of p. 43, so as to make everybody (or no one) happy.



Exercise 3.3

A GPS satellite has, at a given instant, the following position and momentum:

$$\mathbf{r} = [1.4 \times 10^{+7}, 1.6 \times 10^{+7}, 0] \text{ m}$$

$$\mathbf{P} = [-3.1 \times 10^{+6}, 3.4 \times 10^{+6}, 0] \text{ N} \cdot \text{s}$$

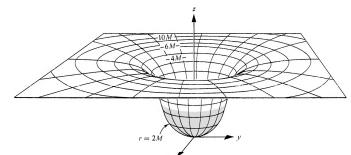
Assuming that the satellite's volume can be considered as small for some purpose, calculate the satellite's angular momentum.

What are energy, momentum, angular momentum?

From the discussions and formulae above, it seems that energy-mass, momentum, angular momentum are quite closely related to one another. For all three, the amount in a volume or through a surface is undefined unless we specify a coordinate system. And we shall see later that all three satisfy balance laws but not necessarily conservation laws.

Relativity Theory indeed shows that energy, momentum, angular momentum are different aspects of one single geometric object, called *energy-momentum tensor*. They are like its "shadows", that we can observe by looking at it from different points of view in time and space. This is also why their values get intermixed if we change our system of coordinates.

General Relativity gives a new meaning to these quantities: they are *particular curvatures of spacetime*. They express how spacetime is curved in different directions. So whenever we measure, say, the energy or the momentum of some object or of some electromagnetic radiation, we are actually measuring how much that object or radiation is curving spacetime in a particular way.



Energy, momentum, angular momentum are measures of particular curvatures of spacetime.

3.7 Entropy

Entropy is a *scalar* quantity, with SI dimension energy/temperature, and measured in units of *joules/kelvins (J/K)*³¹. From this definition it would seem that entropy is derived from temperature. However, although temperature is taken as primitive by the SI, the [definition of temperature](#)³² actually depends on a fixed value of *Boltzmann's constant*³³, which has the dimension of entropy.

Entropy is probably the most difficult quantity to grasp intuitively. Many seemingly intuitive descriptions given in some textbooks are, unfortunately, unhelpful and even misleading. One particularly misleading intuition is that entropy would be a “measure of disorder”. Besides the fact that “disorder” is very vague and subjective, it turns out that some physical phenomena, for example with liquid crystals³⁴, can be considered more “disordered”, and yet have *lower* entropy, than others. See also the example in the side figure. We shall discuss more about such phenomena later on.

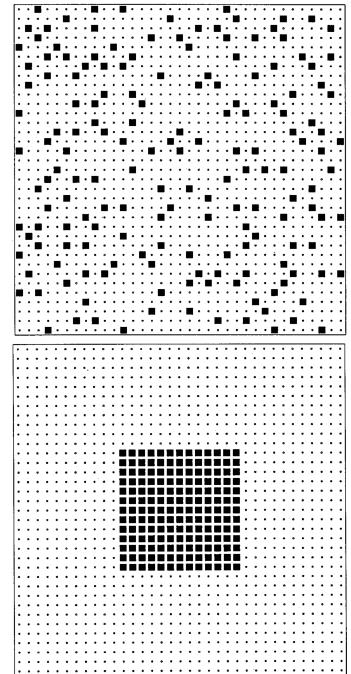
In these notes we shall rely on the idea that *entropy expresses a limit on the flux of energy into matter*. Said in simpler but more imprecise words, entropy is a bound on how fast we can heat up a body. We shall develop this idea further later.

One reason why entropy is difficult to grasp intuitively is that it has very different physical and mathematical aspects depending on the spatial scales and physical theory that we use to describe physical phenomena.

In many “continuum” phenomena, that is, phenomena where the molecular constitution of matter is not visible or not taken into account, entropy is treated as a “stuff-like” quantity similar to energy or electric charge. But there are difficulties also in this case. For some phenomena, for example involving non-elastic materials such as a simple paper clip, it is possible to introduce several entropies having different values – and not just because of a change in measuring scale – all of which can serve their purpose perfectly fine.

In molecular phenomena involving statistical mechanics, on the other hand, entropy is no longer a physical notion, but a *probabilistic* and *statistical* one, related to guesses and inferences that we make about the physical phenomenon. Yet from many points of view it has roles similar to those of the entropy used in continuum phenomena.

We shall see later that the physical laws for entropy have also a different status with respect to the laws for the other six main quantities: they are, so to speak, “laws about laws”.



Microscopic configurations of a lattice gas. Left: a configuration coming from a *low-entropy* state. Right: a configuration coming from a *high-entropy* state (Styer 2000).

Entropy: notation

The amount of entropy in a region is usually denoted with S . We shall see that the flux of entropy is tightly related to heat, and we won't need a special symbol for it.

3.8 Auxiliary quantities

Besides the seven principal quantities, other auxiliary quantities appear in some physical theories. Important examples are **temperature** and **metric**. Most auxiliary quantities don’t have the “stuff-like” properties of the seven principal quantities. For instance, we cannot ask “what’s the total amount of temperature in this region?”. We shall later discuss and use some auxiliary quantities, especially temperature.

The dimensions, units, and scalar or vector character of all quantities mentioned so far are summarized in table 3.1.

| Quantity | SI Dimension | Unit |
|-------------------------|--|--|
| Time | time | <i>second</i> s |
| Length | length | <i>metre</i> m |
| Temperature | temperature | <ikelvin< i=""> K</ikelvin<> |
| Matter | amount of substance | <i>mole</i> mol |
| Electric charge | electric charge | <i>coulomb</i> C |
| Magnetic flux | magnetic flux | <i>weber</i> Wb |
| Energy | energy, mass | <i>joule</i> J, <i>kilogram</i> kg |
| Momentum | force · time, mass · length/time, energy · time/length | N · s, kg · m/s, J · s/m |
| Angular momentum | force · length · time, mass · length ² /time, energy · time | N · m · s, kg · m ² /s, J · s |
| Entropy | energy/temperature | J/K |

Table 3.1 Dimensions and units of the main physical quantities used in these notes.
 Quantities in **boldface** are vectors, the others are scalars

URLs for chapter 3

1. <https://www.youtube.com/watch?v=nYg6jzotiAc&t=893s>
2. <https://doi.org/10.1351/goldbook.A00297>
3. <https://doi.org/10.1351/goldbook.M03980>
4. <https://www.iaea.org/newscenter/news/what-are-radioactive-sources>
5. <https://www.iaea.org/newscenter/news/what-is-nuclear-energy-the-science-of-nuclear-power>
6. <https://www.ciaaw.org/radioactive-elements.htm>
7. <https://pubchem.ncbi.nlm.nih.gov/element/Hydrogen>
8. <https://pubchem.ncbi.nlm.nih.gov/element/Helium>
9. <https://pubchem.ncbi.nlm.nih.gov/element/Lithium>
10. <http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/hadron.html#c6>
11. <http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/lepton.html#c1>
12. <https://www.britannica.com/science/antimatter>
13. <https://www.britannica.com/topic/positron-emission-tomography>
14. <https://pubchem.ncbi.nlm.nih.gov/element/Calcium>
15. <https://www.symmetrymagazine.org/2009/07/23/antimatter-from-bananas>
16. <https://www.quantumdiaries.org/2009/07/21/positrons-from-bananas/>
17. <https://doi.org/10.1351/goldbook.E01923>
18. <https://doi.org/10.1351/goldbook.C01365>
19. <https://doi.org/10.1351/goldbook.A00300>
20. <https://doi.org/10.1351/goldbook.M03684>
21. <https://doi.org/10.1351/goldbook.W06666>
22. <https://doi.org/10.1093/acrefore/9780190871994.013.21>
23. <https://spaceplace.nasa.gov/solar-activity/>
24. <https://doi.org/10.1351/goldbook.K03391>
25. <https://nuclearmuseum.pastperfectonline.com/Archive/716477C1-5E7A-485C-8BE1-857919471563>
26. <https://www.britannica.com/science/nuclear-fission>
27. <https://www.britannica.com/story/atomic-bombing-of-hiroshima>
28. <https://home.cern/tags/13-tev>
29. <https://doi.org/10.4209/aaqr.2019.04.0177>
30. <https://www.britannica.com/science/polymer>
31. <https://doi.org/10.1351/goldbook.C01365>
32. <https://doi.org/10.1351/goldbook.K03374>
33. <https://doi.org/10.1351/goldbook.B00695>
34. <https://www.britannica.com/science/liquid-crystal>

Chapter 4

Fluxes and volume integrals

For the sake of persons of these different types, scientific truth should be presented in different forms, and should be regarded as equally scientific, whether it appears in the robust form and the vivid colouring of a physical illustration, or in the tenuity and paleness of a symbolical expression.

J. Clerk Maxwell 1870

In ch. 3 we said that the main seven quantities – matter, electric charge, magnetic flux, energy, momentum, angular momentum, and entropy – have three common properties related to their measurement:

- (1) We can measure the amount of quantity within a three-dimensional region, at a specific time instant
- (2a) We can measure the amount of quantity flowing through a two-dimensional surface during a time lapse...
- (2b) ... or alternatively we can measure the amount of quantity flowing through a two-dimensional surface *per unit time*, at a particular time instant
- (3) The amount in a volume consisting of separate volumes is equal to the total of the separate amounts. Similarly for the flux through separate surfaces.

Let's give definite names to the measurements (1) and (2b):

volume integral and flux

We can call measurement (1) the **volume integral** of the quantity, although we won't use this term very often.

Most important, we call measurement (2b) the **flux** of the quantity.

These two notions and measurements are very intuitive; that's also why it's convenient to base our physics upon them. In this chapter we straighten some details about their definition and also about our intuition.

4.1 Control volumes and control surfaces

A first question that comes to mind is: how are the volumes and surfaces chosen?

The choice of volume (and therefore of the surface bounding it) for a volume integral, and the choice of surface for a flux are *completely arbitrary*, and they can be *completely imaginary*. Since they are under our control, they are called **control volumes** and **control surfaces**.

For example, consider a classroom and the people in it. In your imagination you can divide the classroom into two halves, say the front and the rear half. You can then ask or measure simply by counting: (1) how many people are, right now, in the rear half; (2) how many people are crossing the imaginary division between the front and rear half during one minute, starting from now.

A control volume and a control surface don't need to be static: they can move and deform.

In the case of a control volume, movement doesn't matter: the volume integral in a control volume *does not depend on the instantaneous motion of the volume*. In fact we can even imagine a control volume that exists for just one time instant.

In the case of a control surface the situation is different. The flux through it *depends also on the motion of the surface*. As a trivial example, consider a glass surface, and a person on one side of it, moving with a high velocity directed towards the surface. Will the person crash on the glass? We can't say for sure. The glass surface could be a glass wall in a building, which is not moving; in this case the person will likely crash on it. Or it could be the windscreen of a car, which is moving together with the person, who's the driver; in this case the person won't crash on it.

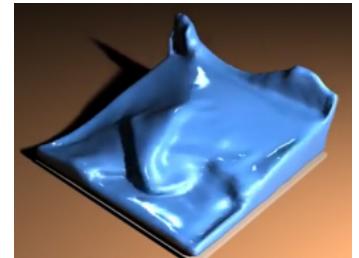


The golf ball is moving leftwards. Will it hit the metal surface? We don't know unless we know how the surface is moving.

So we can't just imagine a surface that exists for one time instant: we need to imagine it for a very short time lapse, and be able to say how it's moving. If someone asks you what's the flux through a control surface at a given instant, but they don't tell you what's the motion of the surface, then the flux is unknown.

The fact that we can choose control volumes and control surfaces arbitrarily gives us a lot of power in solving physics problems and in making predictions. Typically they are chosen so as to simplify the equations that describe the physical situation, simulate physical phenomena in a more precise way, and focus on details of interest.

When we study some solid object, like a football, a rocket, or a planet, we typically choose a control volume that tightly encloses the object. When we study something flowing or moving, such as a fluid material or an electromagnetic field, we typically divide the space of interest into small control volumes and surfaces, constructing a *mesh*; this mesh can even be refined in regions that are of special interest.

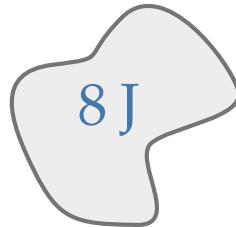


Clever choice of control volume and surfaces allow us to model and predict complex motions of fluids; see also animation¹ (Wojtan et al. 2009)

4.2 Volume integrals: intuition and visualization

Scalar quantities

A volume integral for a scalar quantity, for example energy, can be represented like this:



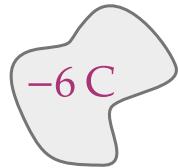
we have eliminated one spatial dimension for simplicity, considering the analogous two-dimensional idea. The volume is in light grey, delimited by a closed darker grey boundary, and we're indicating that the volume integral, that is, the amount of energy within, is **8 J**.

As a visualization device, this representation can be useful. But let's straighten out some of its aspects:

- Recall that this is a snapshot at a given time instant. So there are 8 J of energy in the volume at that instant, but we don't know the situation earlier or later: there could be a different amount of energy, the region

might be at a different position and have a different shape, or it might not even exist.

- Recall that some scalar quantities, such as electric charge and in some situations matter (antimatter), can have negative amounts.
- We must not surmise that the amount of quantity is uniformly distributed within the volume. In fact there could be negative amounts of it in some subvolumes and positive in others. In particular, even if there is a zero amount of quantity in a volume, some subvolumes could have non-zero amounts: some positive and some negative, so that the total is zero.



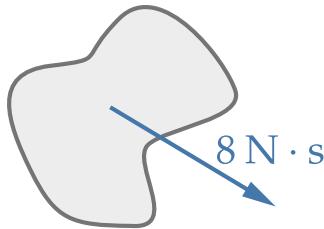
A region with a negative amount of charge

Exercise 4.1

The volume integral of matter in a particular volume is equal to 36 mol. Can we conclude that the volume doesn't contain antimatter?

Vector quantities

A volume integral for a vector quantity, for example momentum, can be represented as follows (we still simplify our visualization to two dimensions):



Momentum is a vector quantity, so the total amount in the volume above is a vector. The picture shows the direction and orientation of this vector, and the magnitude of $8 \text{ N} \cdot \text{s}$ is explicitly reported.

Vector magnitudes and opposite vectors

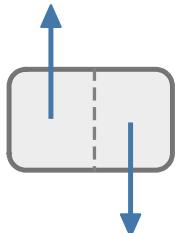
Remember that the *magnitude* of a vector is always positive, and that

$$\overrightarrow{\text{v}} = -1 \cdot \overleftarrow{\text{v}}$$

The visual representation above is useful, if we keep in mind remarks analogous to the scalar case:

- This is a time snapshot.
- The application point of the vector representing the volume integral is unimportant: for instance, it doesn't need to be placed at the centre of the volume. The vector refers to the volume as a whole, not to some specific point within.
- Different subvolumes could have amounts represented by different vectors; only the total vector is represented above.

This last remark will be especially important when we discuss some physical phenomena involving rotation. As an example, look at the side picture: the volume integral for the whole region is *zero*, but its left and right subregions have *non-zero and opposite* volume integrals.

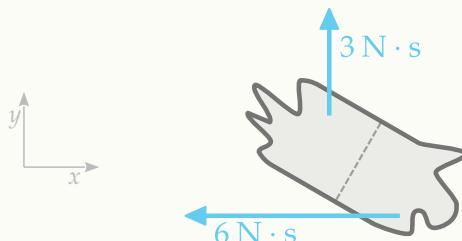


The whole region has zero volume integral. The left and right subregions have non-zero and opposite volume integrals.

Exercise 4.2

Remember the third property (p. 51) of our seven main quantities: The amount in a volume consisting of separate volumes is equal to the total of the separate amounts.

We have a region consisting of two subregions; the amounts of momentum in each subregion are shown below.



1. Write the total momentum in each subregion in component form, (P_x, P_y) , according to the coordinate system shown.
2. Calculate the momentum in the whole region; represent it graphically as vector and write it in component form.

Adding vectors in General Relativity

We are used to the idea of adding vectors placed at different points in space: we only have to first move each – keeping it parallel to itself – to a common point, and then add them all at that point with the usual rule.

This operation *cannot* be done in General Relativity: the notion of parallelism doesn't apply anymore in a simple way, owing to the curvature

of spacetime. The addition would lead to different results depending on how we transported the vectors. So in General Relativity we can only sum vectors that are placed at the same spacetime point.

How can this operation be possible in Newtonian mechanics and in practical applications, then? After all, General Relativity surely applies here! The answer is that the discrepancies of vector transportation are small enough in the neighbourhood of the Earth, as the curvature of spacetime is very small here.

4.3 Fluxes of scalar quantities: intuition and visualization

The direction, reckoning, and representation of scalar fluxes

Earlier we took the intuitive example of a flow of people through an open door; we might ask, for instance, how many people crossed the door in a minute. But one more detail about this flow is important: in which direction did the persons cross the door? For example, if the door leads to a classroom, we may need to know whether the people who crossed the door got in or out, so as to know if there are seats left in the classroom.

In order to do this we can: 1. Assign a crossing direction to the door, calling for instance ‘positive’ the direction from outside to inside the classroom. 2. Count as ‘positive’ each person who crosses the door in the positive direction, and count as ‘negative’ each person who crosses the door in the opposite direction. The total tells us the *net* number of people who *entered* the classroom. If the total is positive, then more people got in than out; if the total is negative, then more people got out than in.

One important aspect of this example and terminology is the following symmetry:

- What crossing direction is called ‘positive’ is fully arbitrary, just a matter of agreement.
- If we decide to call ‘positive’ the other crossing direction, then the total will change sign. But the physical situation is of course still the same.

Therefore the sentences “+5 persons entered the room” and “–5 persons exited the room” are saying exactly the same thing. This somewhat trivial fact about fluxes will, later on, turn out to be connected with a famous law. So let’s remember:

 Every flux is equivalent to a flux of opposite amount in the opposite direction

A flux in a particular surface-crossing direction is equivalent to a flux of *opposite sign* in the *opposite* crossing direction.

"LEX III. Actioni contrariam semper & æqualem esse reactionem: sive corporum duorum actiones in se mutuo semper esse æquales & in partes contrarias dirigi."

Newton 1726

Another important aspect of the example above is that if we're told “ -5 persons exited the room”, we don't know how exactly this happened: it could be that 5 persons got into the room during that minute; or that 10 persons got in and 5 got out; and so on. This is why we call ‘ -5 ’ the *net* amount.

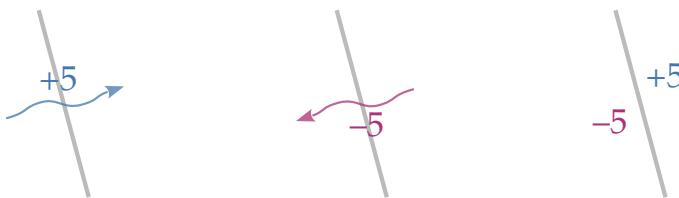
Finally, consider a similar example but with a quantity that can ordinarily also be negative, such as electric charge. Call ‘positive’ the crossing direction from outside to inside the room. If we're told that a net charge amount of -5 crossed the door in the positive direction in one minute, then this could have happened in several ways:

- a charge of -5 was brought into the room
- a charge of $+5$ was brought out of the room
- a charge of -2 was brought into the room during the first 30 s, and a charge of $+3$ was brought out in the remaining 30 s
- a charge of -2 was brought into the room during the whole minute, and a charge of $+3$ was brought out at the same time
- ... and many other possible combinations.

Indeed, ordinary electricity was thought for some time to be associated with movements of negative *and* positive charges in opposite directions.

The purpose of the examples and scenarios above is just to make you aware of some aspects of what we shall call “flux”, which are trivial but important when considering fluxes of physical quantities.

How can we graphically represent the flux of a quantity, in such a way as to take care of these aspects? Consider these three representations:



The grey straight line represents a surface (simplified to two dimensions) through which we're measuring a flux. In the first and second picture, the wavy arrows represent arbitrary crossing directions called ‘positive’.

"Fechner [in 1845] supposed every current to consist in a streaming of electric charges, the vitreous charges travelling in one direction, and the resinous charges, equal to them in magnitude and number, travelling in the opposite direction with equal velocity." Whittaker 1951

The blue arrow represents that we're calling 'positive' the left-to-right direction; the red arrow represents that we're calling 'positive' the right-to-left direction. The signed number represents the net amount according to the positive direction – so the first and second picture represent the same thing. One possible drawback of these two pictures is that they may suggest that a given amount is actually moving from left to right or vice versa; but we have seen that in general we don't know this. The third picture tries to avoid this misleading suggestion by not showing any arrows; it is meant to represent that on the left side the amount of quantity has changed by -5 , and on the right side by $+5$.

The third representation above has one more advantage. Remember from § 4.1 that a flux through a surface may occur because the surface itself is moving. The wavy arrows in the first two representations above may misleadingly suggest that some amount of quantity is "moving" in their direction, or that the surface itself is moving in that direction. The third, arrow-less representation is less misleading.

In these notes we shall settle on the third representation above, but feel free to use the one you prefer – as long as you are aware of all the important aspects of a flux.

What a flux does and doesn't tell

A flux tells us the net change in the amounts of a quantity on the two sides of a surface. These two amounts are equal in magnitude but have opposite signs.

A flux does *not* tell us:

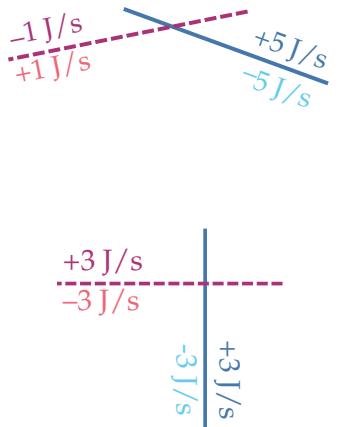
- whether the quantity "was in motion"
- ... and if the quantity was in motion, what was its sign
- whether the surface was in motion

How does a scalar flux change, if we change the surface?

We know that the flux through a surface consisting of two smaller surfaces is the sum of the fluxes through the smaller surfaces. But what if we consider a *different* surface, maybe intersecting the original one? It's important to keep in mind that a **flux refers to a particular surface, and can be very different if we consider a different surface, even if it's close to the original one.**

Consider for instance the picture on the side. We have two intersecting surfaces (as usual simplified by removing one dimension), both almost horizontal. The energy flux through the **solid blue surface**, in a roughly upward direction, is **+5 J/s**. The energy flux through the **dashed red surface**, again in a roughly upward direction, is instead **-1 J/s**.

As another example, the new side picture shows two similar intersecting surfaces, one fully vertical in **solid blue**, and one fully horizontal in **dashed red**. The energy flux through the first in a rightward direction is **+3 J/s**, and so is the energy flux in an upward direction through the second surface.



Flux units

Remember that the flux of a quantity is defined as an amount of that quantity *per time lapse*. Therefore the physical dimension of the flux is

$$\frac{\text{quantity}}{\text{time}}$$

and its units will be the units of the quantity divided by seconds:

| units of fluxes of the four scalar quantities | | | | |
|---|--------|-----------------|--------|-----------|
| quantity | matter | electric charge | energy | entropy |
| flux units | mol/s | C/s | J/s | J/(K · s) |
| equivalent units | | ampere A | watt W | |

After our discussion about the peculiarity of fluxes it's quite easy to work with the fluxes of the four main scalar quantities: matter, electric charge, energy, entropy. Let us add some reminders and remarks about the fluxes of matter and energy.

Matter flux

Remember that *antimatter* “counts as -1 ” for calculating amounts of matter. If 1 mol of positrons (anti-electrons) crosses a surface from left to right in 1 s, the left-to-right flux equals -1 mol/s – note the minus sign. The fact that antimatter is given special names can lead to ambiguities. For instance, if someone asks “what's the left-to-right flux of *positrons*?”, maybe we should answer “1 mol/s”, since the question concerned specifically positrons. It's somewhat like asking “what's the flux of *negative* electric

charge?". In these ambiguous situations is best to add some explanatory words.

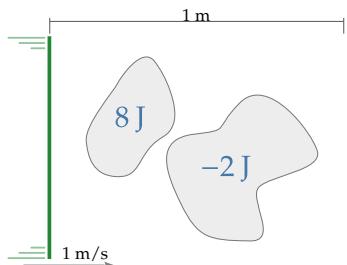
Energy flux

In § 3.4 we mentioned that energy flux can be categorized into different kinds, depending on whether there are fluxes of other quantities through the same surface. We study the exact definitions and formulae later on. The total flux is given by the sum of all these kinds. For instance, through a horizontal surface we can have a downward energy flux of 3 J/s as *heat*, and a downward flux of -1 J/s as *work*. The total downward flux is then 2 J/s. The energy flux that you will calculate in the fourth exercise below is called *energy convection*.

Exercise 4.3

For each question, answer in an *unambiguous* way and sketch a picture representing the flux.

1. The two sides of a particular surface are called 'up' and 'down'. During 0.2 s, an energy of +3 J flows from the up-side to the down-side, and an energy of -4 J flows from the down-side to the up-side. How much is the flux of energy through the surface?
2. Through the same surface, at a later time, 2 mol of neutrons flow from the up- to the down- side in 0.01 s, and 2 mol of neutrons flow from the down- to the up-side during the same time. How much is the flux of matter through the surface?
3. The two sides of a surface are called 'in' and 'out'. During 0.01 s there is a flow of 1000 electrons from the in-side to the out-side, and also a flow of 1000 *positrons* (anti-electrons) in the same direction. How much is the flux of matter through the surface?
4. The side picture shows a **surface** moving from left to right at a (constant) velocity of 1 m/s. The space to its right has two static regions with some amount of **energy** as shown (there's no energy behind to the left of the surface). How much is the flux of energy through the surface in 1 s?



4.4 Fluxes of vector quantities: intuition and visualization

The direction, reckoning, and representation of vector fluxes

The flux of a vector quantity is also a vector, because it is given by an amount of that quantity, which is a vector, divided by time, which is a scalar.

The discussion of § 4.3 about the arbitrary choice of a crossing direction and the minus signs that appear when we reverse it also applies, in an analogous way, to vector fluxes. Also the discussion about the minus sign remains the same; we must only remember that a minus sign changes the sense or orientation of a vector:

$$\overleftarrow{} = -1 \cdot \overrightarrow{}$$

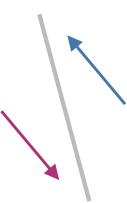
Take for instance a horizontal surface. Suppose that we call ‘positive’ the *upward* crossing direction, and that with this convection the flux is represented by the vector



Then if we decide to call ‘positive’ the *downward* crossing direction instead, the same flux is represented by the opposite vector



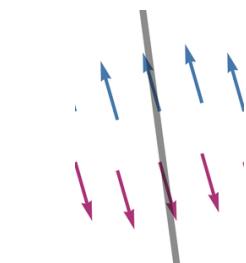
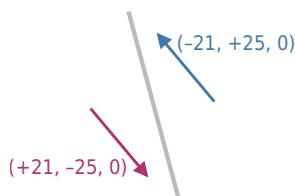
The situation in the case of vector fluxes can require a little more thinking and attention, because it’s easy to get confused between the crossing direction for the surface, and the direction suggested by the vector flux – which are completely separate things. For this reason a good graphical representation of a vector flux is this, analogous to the third one for scalar fluxes on p. 57:



The grey straight line represents a surface through which we're measuring the flux of a vector quantity; on the left side of the surface, the amount of quantity has changed by  and on the right side by . Obviously these two changes have equal magnitude but are opposite.

As a mental image of a vector flux, you can imagine a flow of vectors crossing the surface in one direction, together with a flow of opposite arrows crossing the surface in the opposite direction. See this link² as an example.

Since a vector is represented by three numbers (which can be positive or negative), a vector flux can also be interpreted as the collection of three distinct numerical fluxes:



Imagine the blue upper arrows moving from left to right, and the red lower arrows moving from right to left. Animated version here³.

the picture says that the x -component of the quantity has changed on the right side of the surface by -21 , and on the left side by $+21$; the y -component has changed on the right side by $+25$, and on the left side by -25 ; and the z -component hasn't changed on either side.

Another aspect of vector fluxes that we must try not to get confused about is the application point of the vector representing the flux (the base point of the arrow). This is a time snapshot. Just like in the case of vector volume integrals (p. 55), **the application point of the vector representing the flux is unimportant. The vector refers to one side of the surface as a whole.**

Exercise 4.4

A horizontal surface is given, and there is a flux of a vector quantity through it (for the moment we neglect units):

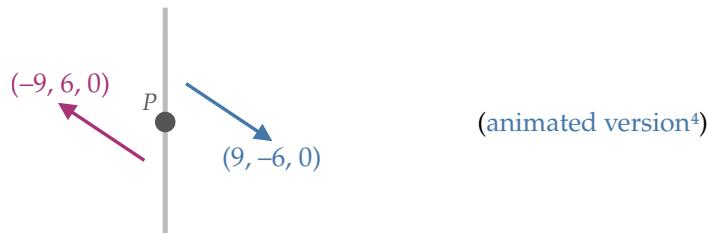
1. If we take the *downward* crossing direction as 'positive', the flux xyz -components are $(5, 5, 0)$. Represent this flux graphically, in the way discussed in the present section. Use the coordinate system  where y points upward.
2. Taking the same crossing direction, represent graphically the flux $(0, -2, 0)$ instead.

3. Taking the same downward crossing direction, we are now told that there is a flux with components $(1, -2, 3)$. What are the components of this flux if we take the *upward* crossing direction as positive?

How does a scalar flux change, if we change the surface?

We saw that the flux of a scalar quantity can be very different if we take a slightly different surface. The same is true of the flux of a vector quantity: in particular, **the vectors representing the fluxes through two slightly different surfaces can point in completely different directions**.

Here is an example. Take a fixed point P . Now take a small vertical surface passing through P . The flux of a vector quantity (momentum for example) through this quantity can be as in this picture:



it has components $(9, -6, 0)$, with magnitude around 10.8, if we take the rightward crossing direction as positive.

Now forget about that surface, and take instead a small horizontal surface passing through the same point P . The flux – of the same quantity – through this surface can be as in this picture:



it has components $(-6, -3, 0)$, with magnitude around 6.7, if we take the upward crossing direction as positive.

Clearly the vectors representing the fluxes through these two surfaces are different: they point in different directions, and have even different magnitudes.

If we have a vector flux through a particular surface, and we're asked about the flux through a different surface, we might be tempted to "move"

through the new surface the vectors representing the flux through the old one. *We must fully resist this temptation.* It only leads to mistakes.

Each surface has a unique flux

A flux, scalar or vector, through a particular surface, in general doesn't tell you anything about the flux though another surface, even if the other surface is only slightly different from the first one.

4.5 Flux of momentum: force

We already mentioned in § 3.5 that **flux of momentum is what we call ‘force’**. Owing to the importance of the notion of *force* in the many branches of physics which rely on Newtonian mechanics, we must discuss this connection in depth. This connection, as well as the connection to Newton's laws, will become even clearer when we discuss the balance of momentum in § 8.

Units

In § 3.5 we said that momentum can be measured in newton-seconds, ‘N · s’. Its flux, being a momentum per unit time, is therefore measured in newtons, ‘N’.

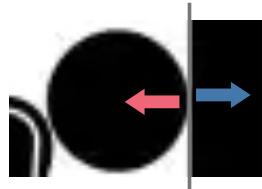
Representing and visualizing force as flux of momentum

The notion of force is very intuitive. We associate it to the sensations that we feel in our skin, flesh, and even bones when, for instance, we push against a wall, twist a door knob, push backwards on the ground with our feet to run, or other similar actions. This force is typically represented by a vector, having the direction and orientation of the “push” or “pull”, and magnitude expressing its intensity. *Such a force vector is exactly the vector expressing the flux of momentum.* The two are the same. A force can therefore also be visualized as a flow of momentum. This mental representation can be illuminating in some physical problems.

As a concrete example, imagine a person pushing against a wall. In terms of force, we say that *the person is exerting a force on the wall, and the force vector on the wall has a person→wall orientation*. This is usually depicted, for example, like this:

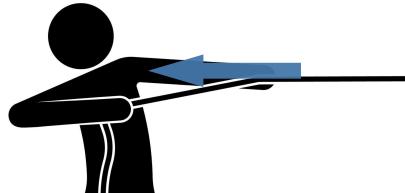


In terms of momentum flux, we imagine a surface separating the person and the wall. If we take a person→wall crossing direction, the flux vector also has a person→wall orientation. Because of the symmetry of flux, if we take a wall→person crossing direction instead, then the flux vector also changes direction. This momentum flux can be depicted (zooming in) like this:



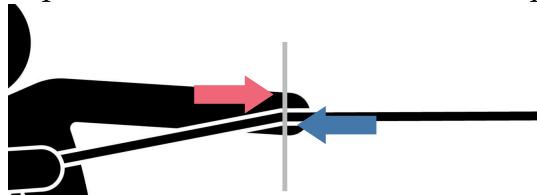
The picture says: on the side of the wall, momentum is changing by an amount having [person→wall orientation](#); and, by symmetry of the flux, on the side of the person's head, momentum is changing by an amount having opposite, [wall→person orientation](#).

Let's take an example involving pulling instead of pushing. Imagine a person pulling a rope fastened somewhere. In terms of force we say that *the person is exerting a force on the rope, and the force vector on the rope has a rope→person orientation*. This can be depicted like this:



In terms of momentum flux, we imagine a vertical surface between the person's hand and the rope. If we take a hand→rope crossing direction, the flux vector has a rope→hand orientation – note the difference from the previous example. Because of the symmetry of flux, if we take a

rope→hand crossing direction instead, then the flux vector has opposite direction: hand→rope. This momentum flux can be depicted like this:



The picture says: on the side of the rope, momentum is changing by an amount having [rope→person orientation](#); and, by symmetry of the flux, on the side of the hand, momentum is changing by an amount having opposite, [hand→rope orientation](#).

A final example illustrates a situation in between the previous two. Consider the foot of a running person, as it pushes on the ground. In terms of momentum flux, we imagine a horizontal surface between the runner's foot and the ground. If we take a downward crossing direction, the flux vector is oriented towards the back of the foot. Because of the symmetry of flux, if we take an upward crossing direction, then the flux vector has opposite direction, towards the front of the foot, the same as the running direction. This momentum flux can be depicted like this:



The picture says: on the side of the ground, momentum is changing by an amount having [leftward orientation](#); and, by symmetry of the flux, on the side of foot, momentum is changing by an amount having [rightward orientation](#).

Newton's Third Law!

From the examples above, we see that thinking of force as momentum flux automatically leads to *Newton's third law*: if one side is gaining/losing momentum with a given orientation, by symmetry the other side is gaining/losing momentum with the opposite orientation. So if one side is experiencing a force with a given orientation, the other side must be experiencing a force with the opposite orientation.

"LAW III. To every action there is always opposed an equal reaction: or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts."

(Original Latin on p. 57)
Newton 1726

We see that Newton's third law is the expression of the symmetry of fluxes *for the specific case of the flux of momentum (force)*. But we realize that this property is more general: it applies not only to force, but also to the flux of all other quantities, even scalar ones.

Exercise 4.5

Using your intuition, try to guess the various momentum fluxes (except their magnitudes) between this person and the walls:



(Buster Keaton in '*The Electric House*'⁶)

4.6 Pressure, tension, shear force

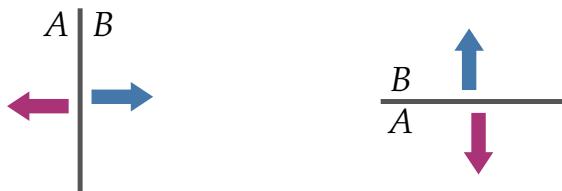
The examples of the previous section demonstrated a variety of possible orientations of the momentum-flux vector with respect to the surface through which it occurs. Obviously all orientations are possible. Special names are given, however, to a momentum flux having three specific orientations: *pressure*, *tension*, and *shear force*.

Take a surface and call its sides *A* and *B*. Now consider a flux of momentum according to the crossing direction $A \rightarrow B$.

Pressure

If the momentum-flux from *A* to *B* is a vector also oriented from *A* to *B*, then we call the momentum flux a **pressure**, or *compressive momentum flux*

or *compressive force*. Here are two examples of pressure:



[\(animated version⁷\)](#)

Pressure is the kind of momentum flux that we exert when we *push* on an object, and that air exerts on all objects it surrounds.

Tension

If the momentum-flux [from A to B](#) is a vector oriented from B to A, then we call the momentum flux a **tension**, or *tensile momentum flux* or *tensile force*. Here are two examples of tension:



[\(animated version⁸\)](#)

Tension is the kind of momentum flux that we experience in our bones when we *pull* an object, and that occurs in any section of a stretched rubber band.

Shear force

If the momentum-flux [from A to B](#) is a vector oriented transversally, along the surface itself, then we call the momentum flux a **shear force**, or *shearing momentum flux*. Here are two examples of shear force:



[\(animated version⁹\)](#)

Shear force is the kind of momentum flux that we experience under our feet when we walk or run, and that occurs between a car's wheels and the ground.

In general, a momentum flux won't have any of the three special directions above, but rather a combination of them.

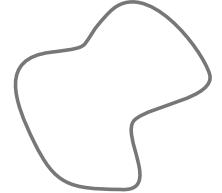
Exercise 4.6

Using your intuition, try to identify the various momentum fluxes that occur in the different parts of a tower crane. Which of the fluxes are (approximately) compressive, tensile, and shearing?



4.7 Closed surfaces, influxes, effluxes

We shall often consider **closed** surfaces, that is, surfaces that don't have a rim or border or holes, like the surface of a sphere or of a cube. A closed surface delimits a specific three-dimensional volume, and we can naturally speak of its **interior** and its **exterior**. An example (simplified by removing one dimension as usual) is the surface in the side picture.



We can give two crossing directions to a closed surface: *inward*, from exterior to interior; or *outward*, from interior to exterior. A flux through the surface is usually called **influx** if we are considering the inward crossing direction, and **efflux** or **outflux** if we are considering the outward crossing direction. Obviously, by the symmetry of fluxes,

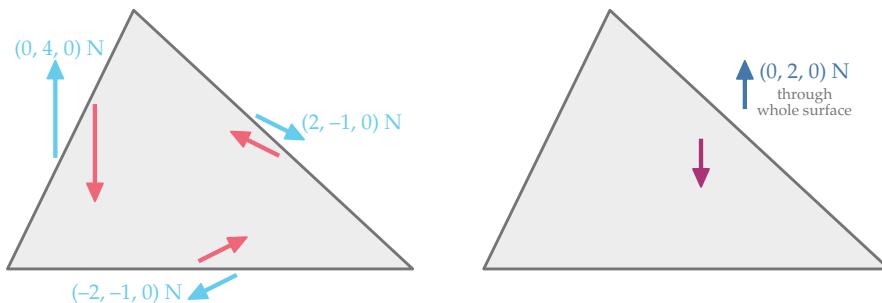
$$\text{influx} \equiv -\text{efflux} \quad \text{efflux} \equiv -\text{influx}$$

The influx and efflux are fluxes *through the whole surface*. Consider for instance these fluxes of energy and of momentum:



On the left we have an efflux of $+5 \text{ J/s}$ on the whole outer side of the surface. We don't know whether these 5 J/s are being evenly distributed all around the surface, or just at particular spots of it. On the right we have an efflux of 8 N , with a vector pointing approximately rightward. Again we don't know what are the flux vectors of individual pieces of the surface: the vector in the picture is just their total.

Let's see another example of this fact. The picture below on the left show the **outward fluxes** through three parts of a closed surface; the picture on the right shows the **total efflux** through the same surface:

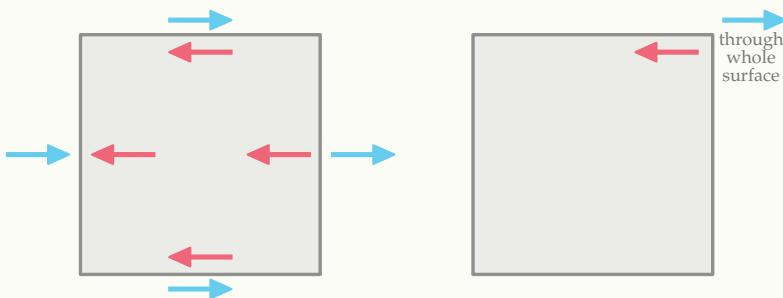


The individual fluxes and the total efflux are consistent, since

$$(2, -1, 0) \text{ N} + (0, 4, 0) \text{ N} + (-2, -1, 0) \text{ N} = (0, 2, 0) \text{ N}$$

Exercise 4.7

1. Are the four partial fluxes shown on the left consistent with the total efflux shown on the right? Why?:



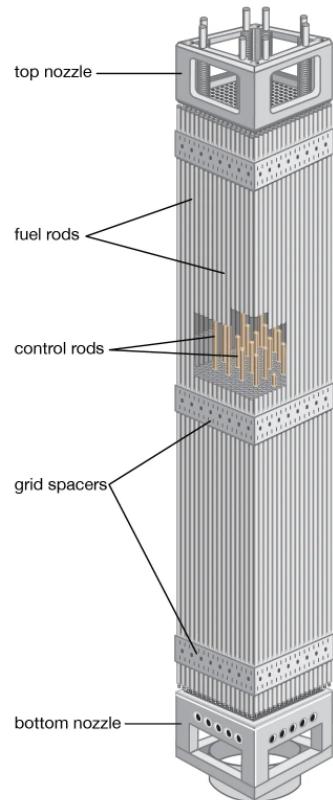
2. Take an imaginary cylindrical surface enclosing one control rod¹⁰ in a nuclear-fission reactor¹¹ (side picture). Let's say that in a reactor there are 20 such rods. Approximately 5×10^{19} neutrons are liberated in a second in the whole reactor by the fission fuel, but $2/3$ of these are *absorbed* by the control rods.

How much, on average, is the efflux of neutrons (matter) through the surface of one control rod?

Express the result first in neutrons/s, and then in mol/s, using the Avogadro constant

$$N_A = 6.022\,140\,76 \times 10^{23} \text{ particles/mol}.$$

Be careful about the signs!



© 2013 Encyclopædia Britannica, Inc.

4.8 Integrated fluxes

A flux is defined as the amount of a quantity crossing a surface in a short time lapse Δt , divided by that time lapse. Denoting the flux by, say, J , this definition also means that the amount of quantity crossing the surface in a short time Δt is equal to $J \Delta t$.

Take now a surface that exists between two time instants t_1 and t_2 ; during this time lapse it could also be moving and changing shape. Choose a crossing direction through the surface. At each intermediate time instant t we can then measure the flux of a quantity crossing the surface in that direction, at that instant; denote it by $J(t)$.

The total amount of quantity that crosses the surface between times t_1 and t_2 can be found by integrating $J(t)$. That is, we divide the time interval into very short time lapses of length Δt ; for each short time lapse we know that the amount that crosses the surface is $J(t) \Delta t$. The total is obtained by adding these small amounts. As Δt is considered shorter and shorter, this sum is by definition an integral:

Integrated flux of a quantity through a surface between two times

The total amount of quantity crossing a surface (relative to a specified crossing direction) between times t_1 and t_2 is given by

$$\int_{t_1}^{t_2} J(t) dt . \quad (4.1)$$

The meaning of the integral above should be clear for any scalar quantity, for which the flux is also a scalar. In the case of a vector quantity, for instance momentum, the flux is also a vector, represented by three components. The integral of a vector is obtained by calculating the integral for each component, obtaining three results, which are the components of a new vector. Geometrically this corresponds to summing a large number of very small vectors.

Take the case of momentum, whose flux (force) we denote $\mathbf{F} = (F_x, F_y, F_z)$. The integral of this flux is then

$$\int_{t_1}^{t_2} \mathbf{F}(t) dt := \left(\int_{t_1}^{t_2} F_x(t) dt , \int_{t_1}^{t_2} F_y(t) dt , \int_{t_1}^{t_2} F_z(t) dt \right) . \quad (4.2)$$

! **The integrated flux can be zero even with non-zero flux**

The result of the integral defining the integrated flux can be zero. This means that no *net* amount of quantity crossed the surface between t_1 and t_2 . Yet the flux $J(t)$ can be non-zero, even at all times; of course it needs to be positive at some times, and negative at others.

As a simple example, consider a room's door. During one minute, three people enter through the door; during the next minute, three people (not necessarily the same) exit through the door. The total flow of people is zero, but the flux was non-zero during the first minute, and non-zero during the second minute.

Exercise 4.8

1. What is the physical dimension of the integrated flux of a quantity?
2. Suppose that we calculate the integral above for a particular surface in the case of matter, finding a total of $\int_{t_1}^{t_2} J(t) dt = 7 \text{ mol}$. Now we change our mind and choose the opposite crossing direction for that surface. How does the result above change?

4.9 Fluxes and velocities

The idea of flux naturally evokes the idea of movement, and therefore of *velocity*. Is there a relationship between flux and velocity?

The answer is yes: velocity is in fact essentially defined from a flux and a volume integral. Consider for example how we measure the velocity of an object: we are actually keeping track of a flow of matter from one region of space to another.

The rigorous definition of velocity from flux is somewhat involved, so here we'll just see a simplified and approximate example of how such a definition work. In any case, in the following chapters you won't need to remember any particular mathematical definition; just keep in mind that there's a tight connection between velocity, flux, and volume integral.

Take an extensive scalar quantity, like matter, charge, energy, or entropy. For concreteness let's think of matter. Take a coordinate system (t, x, y, z) and a point in space at some specific time. Around this point choose a

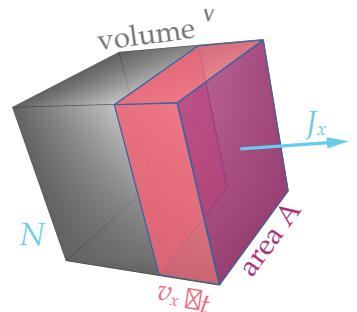
very small cuboid region, delimited by six small rectangular surfaces: two parallel to the yz -coordinate plane, two to the zx one, and two to the xy one. The cuboid has volume V , and the amount of matter (volume integral) in it is N . The flux through either one of the two yz surfaces, crossing in the positive- x direction, is J_x , and either surface has area A . Note that if the cuboid is small enough, the flux through either of the two yz surfaces, and their area, is approximately the same.

The x -component of the velocity of matter in this region, v_x , is defined as

$$v_x := \frac{J_x/A}{N/V} \quad (4.3)$$

with analogous definitions for the y - and z -components. This definition is valid for small velocities and weak gravitational fields.

The velocity $\mathbf{v} = (v_x, v_y, v_z)$ so defined has the following intuitive property. If you choose *any* very small surface centred at this point, and move it with this velocity (in the direction specified by the velocity), then the matter flux through it is zero. This reflects the intuitive understanding that if a surface is “moving” together with the matter, at the same speed, then we shouldn’t observe any flux through it.



A region with a negative amount of charge

Velocities of objects in general relativity

One consequence of the relationship between velocities and fluxes is that we can define such a velocity for any extensive quantity. So we have a velocity matter from the flux of matter, but also a “velocity of energy” from the flux of energy.

In Newtonian approximation these two velocities are equal, so we do not need to distinguish them. In situations where the Newtonian approximation is not valid, on the other hand, we have to take into account the “velocity of matter” and the “velocity of energy” separately. This difference is important for instance in the study of plasma in stars and in numerical general relativity. There is an ongoing discussion as to which of the two velocities is more convenient to use; see for instance Kandus & Tsagas 2008, especially the section *Eckart frame versus Landau frame*, which refers to the choice between these two velocities.

URLs for chapter 4

1. <http://www.youtube.com/watch?v=M5xnAdVPbgQ>
2. <https://pglpm.github.io/7wonders/media/vectorfluxanimsin.webp>
3. <https://pglpm.github.io/7wonders/media/vectorfluxanimsin.webp>
4. <https://pglpm.github.io/7wonders/media/skewflux1.webp>
5. <https://pglpm.github.io/7wonders/media/skewflux2.webp>
6. <https://www.imdb.com/title/tt0013099/>
7. <https://pglpm.github.io/7wonders/media/pressure.webp>
8. <https://pglpm.github.io/7wonders/media/tension.webp>
9. <https://pglpm.github.io/7wonders/media/shearforce.webp>
10. https://energyeducation.ca/encyclopedia/Control_rod
11. <https://www.britannica.com/technology/nuclear-reactor>

Chapter 5

Physical laws

Every branch of physical science is based on two sets of fundamental equations. The first set is that of basic laws of physics, which are postulated to hold valid for all bodies under all conceivable circumstances [...]. The second set of fundamental equations are the constitutive equations: these are relationships which are not supposed to hold for all bodies, but only to describe the behavior of some restricted class of bodies, or possibly of a larger class of bodies for a more restricted class of phenomena.

G. Astarita 1990

5.1 Two kinds of laws

In ch. 3 we introduced seven main quantities in terms of which we describe physical phenomena. How do we describe the patterns, regularities, and diversities that we observe in these phenomena? This is the role of *physical laws*. Different physics formalisms express physical laws in different ways, even if the actual physical phenomenon is exactly the same. In the formalism that we are using we distinguish between two kinds of physical laws:

Main distinction between two kinds of physical laws

- **Fundamental laws** represent universal patterns that we observe everywhere. We shall see that there is a set of seven fundamental laws that apply on every known scale of space and time, and also across different physical theories.

- **constitutive relations** only apply to particular phenomena, only on particular time & space scales, only in particular theories. They express the diversity that we observe around us, and different kinds of approximations that we use in describing physical phenomena.

This distinction it's important. The fundamental laws can be applied with some confidence to every new phenomenon we observe, and often allow us to make predictions of at least a qualitative character. We only modify them as a last resort (which, so far, has been rarely or never necessary). We have a lot of freedom in modifying constitutive relations instead, and in proposing new ones to account for newly observed physical phenomena.

As far as we know there is a very small number of fundamental laws and, remarkably, *all known fundamental laws are balance laws*.

Balance and conservation laws

Balance laws, or simply *balances*, have this name because they express a sort of trade-off or “budget” about physical quantities.

Conservation laws are a special and important kind of balance laws.

In this chapter we shall see how to mathematically express balance and conservation laws, as well as some kinds of constitutive relations.

5.2 Conservation laws

Conservation laws tightly connect the *volume integrals* and *fluxes* discussed in ch. 4. They are intuitively and visually easy to grasp. The archetype is as follows:

Consider a *closed* (§ 4.7) control surface at a (coordinate) time t_1 , which delimits a control volume. A spherical surface, delimiting a ball-shaped region, is an example. Actually the closed control surface doesn't need to be connected: it could consist of several separate closed control surfaces, each delimiting a different volume.

Starting from time t_1 , imagine the closed control surface moving and deforming through time, changing shape and size but remaining closed – no holes or cuts – until a later time t_2 . At this final time we have a final closed control surface that can be different from the initial one, and at a different position. If we wanted, we could start with a spherical surface

delimiting a region the size of a tennis ball, and end with a cubical surface delimiting a region the size of the Sun. The temporal sequence of closed control surfaces delimits a sequence of control volumes; see the sequence on the right as an example.

Now consider any one of the seven primitive quantities of § 3.1, except magnetic flux. Recall that these quantities are *extensive*: we can measure the amount of this quantity in any control volume, and its flux across any control surface. For concreteness let's take *matter*, for which we denote volume integrals by N and flux by J . We make the following measurements:

- $N(t_1)$: total amount of matter contained within the control volume at time t_1 .
- $J(t)$: *influx* (flux crossing inwards, § 4.7) of matter (and antimatter!) through the closed surface at any particular time t between t_1 and t_2 .
- $N(t_2)$: total amount of matter contained within the control volume at time t_2 .

Each of the three amounts above is, in this example, a scalar that can be positive, negative, or zero. For instance we could have:

$$t_1 = 00:00:00 \quad t_2 = 00:00:18$$

$$N(t_1) = 3.50 \text{ mol} \quad J(t) = -0.06 t \text{ mol/s}^2 \quad N(t_2) = -6.22 \text{ mol}$$

note that $J(t)$ is given as a function of the time t .

With this setup we can introduce two kinds of balance laws.

Conservation law in integral form

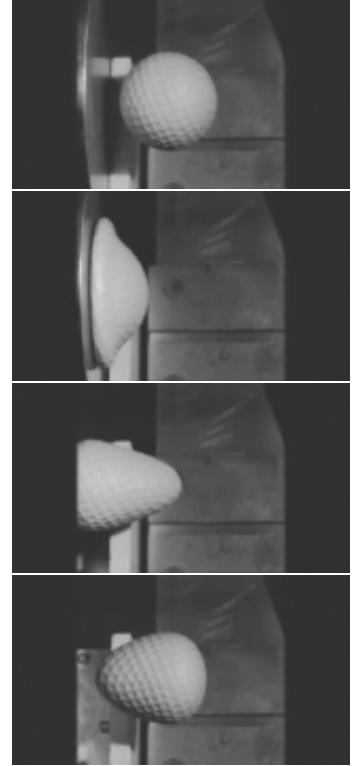
An extensive quantity with volume integral $N(t)$ and influx $J(t)$ is said to satisfy a **conservation law**, or to be **conserved**, if the following equality holds *always*:

$$N(t_2) - N(t_1) - \int_{t_1}^{t_2} J(t) dt = 0 \quad (5.1)$$

The meaning of the equation above is intuitive, especially if we rewrite it like this:

$$N(t_2) = N(t_1) + \int_{t_1}^{t_2} J(t) dt$$

This expression says that the amount of quantity in the final control volume, $N(t_2)$, is equal to the amount in the initial control volume, $N(t_1)$, plus



Temporal sequence of closed control surfaces, represented by the golf ball's surface (click here for an animated version¹)

the total amount (which could be negative) that flowed in between those times, that is, the integrated flux $\int_{t_1}^{t_2} J(t) dt$. Said otherwise, any amount of quantity that appears in (or disappears from) the control volume, must come in (or go out) through the surface; it can't appear out of nowhere.

A conservation law allows us to make several kinds of predictions or deductions. For example:

- If we know the amount of quantity within the control surface at t_1 and the net amount that crossed the surface between t_1 and t_2 , then we can predict the amount within the control surface at t_2 : this is given by the last equation above.
- If we know the amount of quantity within the control surface at t_1 and the one at t_2 , then we can deduce the net amount that crossed the surface between the times t_1 and t_2 :

$$\int_{t_1}^{t_2} J(t) dt = N(t_2) - N(t_1).$$

- If we know the amount of quantity within the surface at t_2 and the net amount that crossed the surface between t_1 and t_2 , we can deduce the amount initially within the surface, at t_1 :

$$N(t_1) = N(t_2) - \int_{t_1}^{t_2} J(t) dt.$$

These kinds of predictions and deductions are very powerful because we are fully free to decide the shape and motion of the control surface and volume, as well as the times t_1 and t_2 .

We unconsciously use several conservation laws all the time in our everyday life. If a bike tyre is suddenly deflated, we conclude that that air must have got out of it through its surface, which must therefore have a hole or a defective valve; we don't conclude that air "just disappeared".

Note that the discussion so far is equally valid for a scalar or a vector quantity (we shall consider vector quantities more in detail in a moment).

Differential form of conservation laws

The conservation law

$$N(t_2) - N(t_1) - \int_{t_1}^{t_2} J(t) dt = 0$$

is said to be written in *integral form*, because it contains the integrated flux, necessary to calculate the net amount of quantity that crosses the control surface.

Change the symbol for time t_1 simply into t , and take t_2 to come after a *very short* lapse of time Δt after t :

$$t_1 \text{ becomes } t, \quad t_2 \text{ becomes } t + \Delta t$$

Then the integrated flux can be approximated by $\int_t^{t+\Delta t} J(t) dt \approx J(t) \Delta t$ (because the flux $J(t)$ is approximately constant in the short time lapse Δt).

The conservation law can then be approximated as

$$N(t + \Delta t) - N(t) - J(t) \Delta t \approx 0$$

If we divide the whole expression by Δt and consider smaller and smaller Δt , the fraction $\frac{N(t+\Delta t) - N(t)}{\Delta t}$ becomes, by definition, a derivative, and we obtain a new form for the conservation law:



Some mathematical conditions must be satisfied to make these steps; otherwise the derivative below must be carefully defined in a generalized sense.

Conservation law in differential form

A quantity is said to satisfy a **conservation law** if the following equality holds always:

$$\frac{dN(t)}{dt} - J(t) = 0 \quad (5.2)$$

This form says that the rate of change of the volume integral of a quantity must equal *influx* of that quantity.

Both forms, integral and differential, of a conservation law are useful, in different situations. The integral form is obviously useful if we know the total flow during a time lapse but we don't know the flux at each instant. The differential form is often useful for simulating physical processes and make predictions.

Example with a moving control surface

The tube within the tyre of a bicycle has, at a given time, 12.6 mol of air. The bicycle and its tyres moves around, and thirty minutes later the tyre is flat, the inner tube having only 0.5 mol of air.

We consider an imaginary, closed control surface wrapping the tube. The control surface deforms just like the tube deforms. The initial control surface has therefore a tubular shape, but the final control surface doesn't.



We assume that air satisfies a conservation law. From the description above we have

$$t_1 = 0 \text{ s}, \quad t_2 = 1800 \text{ s}, \quad N(t_1) = 12.6 \text{ mol}, \quad N(t_2) = 0.5 \text{ mol}.$$

The conservation law (5.1) allows us to calculate the integrated **influx**:

$$\begin{aligned} \int_{t_1}^{t_2} J(t) dt &= N(t_2) - N(t_1) \\ &= 0.5 \text{ mol} - 12.6 \text{ mol} \\ &= -12.1 \text{ mol} \end{aligned}$$

The negative-signed result says that 12.1 mol of air have crossed the moving control surface, in an outward direction, during thirty minutes. Obviously this happened because the tyre tube must have a hole. Note that this is a physical hole in the physical tube; our control surface is imaginary and doesn't have any holes. It's just a sort of "border checkpoint" where anything can in principle move through; we only keep count of what's crossing, how much, and how fast.

We see that the fact that air satisfies a conservation law allows us to determine the integrated flux through the control surface. But it doesn't allow us to determine the flux $J(t)$ at every individual time between t_1 and t_2 . For example, we don't know if there was a larger flux at the beginning than at the end; the tyre may have gone completely flat just after 60 s; in that case the flux $J(t)$ was zero for $t > 60$ s.



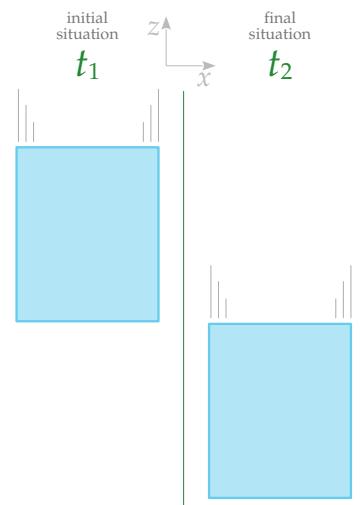
There's a physical hole in the physical tube, but no holes in the imaginary control surface that wraps the tube

Example with a static control surface

In the previous example we chose a closed control surface that moved and deformed with some object of interest (the tyre tube). But we can also choose a different control surface, for instance a static one.

Consider a block of 53.4 mol of ice having cylindrical shape with 1 m diameter and 1.226 m height. It is falling downward owing to gravity (and in a vacuum). At a particular time it occupies a particular position; after 0.5 s it is situated, for an instant, immediately underneath the initial position, as illustrated in the side picture.

Suppose we want to know the net quantity of ice that passed, in a downward direction, across the circular surface separating the initial and final positions. We can calculate this quantity by assuming that ice (as matter) satisfies a conservation law.



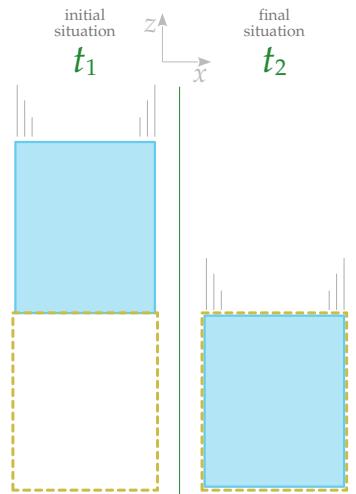
We choose a *static* closed control surface of cylindrical shape. This control surface is located in such a way that it wraps the block of ice at the final time; at the initial time, therefore, the control surface is empty. See the side picture, where the control surface is represented by the **dashed yellow line**. This static control surface consists of three parts: a circular surface at the bottom, a side surface, and a circular surface at the top – which is the one that interests us.

According to the description above we have

$$t_1 = 0 \text{ s}, \quad t_2 = 0.5 \text{ s}, \quad N(t_1) = 0 \text{ mol}, \quad N(t_2) = 53.4 \text{ mol}.$$

From the conservation law (5.1) we calculate the integrated **influx**:

$$\begin{aligned} \int_{t_1}^{t_2} J(t) dt &= N(t_2) - N(t_1) \\ &= 53.4 \text{ mol} - 12.6 \text{ mol} \\ &= 53.4 \text{ mol} \end{aligned}$$



Note that this is the integrated influx through the *whole* cylindrical surface. The conservation law doesn't tell us anything about the integrated influx through *parts* of the surface. In the present case we provide the extra knowledge that the fluxes through the side and bottom surfaces are zero, between t_1 and t_2 . Then, by the extensivity property (p. 31),

$$\begin{aligned} \int_{t_1}^{t_2} J(t) dt &= \int_{t_1}^{t_2} [J_{\text{top}}(t) + J_{\text{side}}(t) + J_{\text{bot}}(t)] dt \\ 53.4 \text{ mol} &= \int_{t_1}^{t_2} [J_{\text{top}}(t) + 0 \text{ mol/s} + 0 \text{ mol/s}] dt \end{aligned}$$

from which we find, as was intuitively clear, that the net amount of ice that crossed the top surface in a downward direction is 53.4 mol.

Exercise 5.1

Solve the following exercises not just by using intuition, but **by explaining step-by-step how you use a conservation law**, in integral or differential form, to obtain the result:

- what is the relevant time interval (if using the integral form)
- how you define the closed control surfaces and their movement, as well as any subdivisions of the surfaces

- what are the volume integral and integrated-flux values you know, and which you want to find

1. For the present exercise we assume that energy satisfies a conservation law.

An apartment's room has two identical electric heaters along a wall. An electric heater can be considered as a piece of surface across which energy flows into the room: the energy is entering (through the electric cable) in the form of electromagnetic energy, and is converted into internal energy (mainly of the room's air) by means of the heater. Suppose that each heater corresponds to an influx of 200 J/s.

The room also has a window, which is the only other part of the room's boundary where energy can flow in or out.

In one hour we measure that the total amount of energy in the room has not changed. How much is the integrated energy *influx* through the window during that time?

2. For the present exercise we assume that Argon² atoms satisfy a conservation law.

A [distillation column](#)³ in a chemical plant has a constant influx of 2 mol/s of Argon atoms at one inlet, and a constant outflux of Argon atoms, possibly at a different rate, at an outlet. An amount of 500 mol of Argon is measured in the chamber at a given time, and one minute later the amount is measured to be 620 mol. How much is the Argon *outflux* (*not* the integrated outflux) at the outlet?



5.3 Balance laws

Our intuitive understanding of a conserved quantity is that it cannot be “created” or “destroyed”. This leads to the definition of a more general law where this condition is not necessarily satisfied.

The setup is just like before: consider an initial and final closed control surfaces and volumes at t_1 , t_2 , and a sequence of closed control surfaces

in between. This time let's take *energy-mass* as a concrete example, with initial volume integral $E(t_1)$, final volume integral $E(t_2)$, and influx $\Phi(t)$.

Balance law

A quantity with volume integral $E(t)$ and influx $\Phi(t)$ is said to satisfy a **balance law**, or to be **balanced**, if the following equality holds *always*:

$$E(t_2) - E(t_1) - \int_{t_1}^{t_2} \Phi(t) dt = \int_{t_1}^{t_2} R(t) dt \quad (5.3)$$

where $R(t)$, called the **supply** or **source**, is the amount of quantity created per unit time within the closed control surface at time t .

If the quantity is a scalar, then $R(t)$ can also be negative, in which case we may call it a *sink* and we say that some of the quantity is being destroyed at time t .

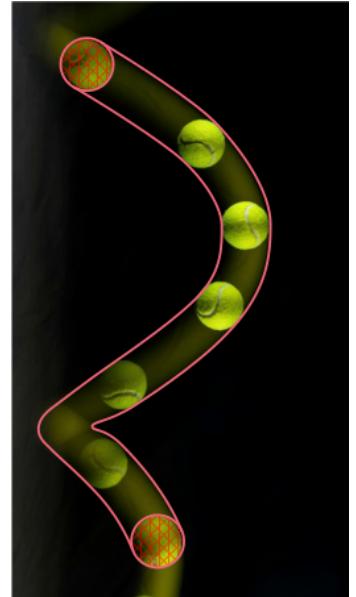
According to the definition above, a conservation law is a *special, powerful case* of a balance law: one for which $R(t) = 0$ *always and everywhere*; that is, there never are sources or sinks of the quantity. It is powerful because the supply $R(t)$ is already known and has an extremely simple value (zero).

For this reason, a balance law is in some respects more trivial than a conservation law. If we measure that $E(t_1) - E(t_2) - \int_{t_1}^{t_2} \Phi(t) dt$ is not zero (so there's no conservation law), we can always say that some amount of quantity must have been created or destroyed within the control volume between t_1 and t_2 . An extensive quantity can therefore always be said to satisfy a balance. A balance is not trivial, however, if we have some specific physical law that tells us in advance how the amount created or destroyed at each instant, $R(t)$, can be calculated.

A conservation law is therefore powerful also because it allows us to predict the amount of quantity in the final volume by making measurements *only on the surface* during the time lapse. A general balance law, instead, requires us to also make measurements at every point *within the volume* during the time lapse.

Supplies are very different from fluxes

One could object: "why don't you just put the flux $\Phi(t)$ and the supply $R(t)$ together, adding the two integrals in equation (5.3)? Wouldn't you



A conservation law only involves what happens within the initial and final control volume (represented by the red hatched disks), and on the sequence of control surfaces in between (represented by the curved red lines), but not what happens within the control volumes in between

get

$$E(t_2) - E(t_1) - \int_{t_1}^{t_2} [\Phi(t) + R(t)] dt = 0$$

which looks like a conservation law?"

Unfortunately the mathematical expression above wouldn't be a conservation law, despite its appearance. The point is this: the flux Φ involves only what's happening *on the control surface*; the supply R , on the other hand, involves what's happening *within the control volumes, at each time*. A conservation law involves only fluxes within the integral. That is, it doesn't require us to know what's happening within the control volumes, except at the initial and final time.

Balance law for vector quantities

Let us now consider an extensive vector quantity like *momentum*. Denote its initial and final volume integrals by $\mathbf{P}(t_1)$ and $\mathbf{P}(t_2)$, and its influx (the force) by $\mathbf{F}(t)$. These are all time-dependent vectors:

$$\mathbf{P}(t) = \begin{bmatrix} P_x(t) \\ P_y(t) \\ P_z(t) \end{bmatrix} \quad \mathbf{F}(t) = \begin{bmatrix} F_x(t) \\ F_y(t) \\ F_z(t) \end{bmatrix}.$$

The balance law has an analogous form:

$$\mathbf{P}(t_2) - \mathbf{P}(t_1) - \int_{t_1}^{t_2} \mathbf{F}(t) dt = \int_{t_1}^{t_2} \mathbf{G}(t) dt \quad (5.4)$$

where the supply

$$\mathbf{G}(t) = \begin{bmatrix} G_x(t) \\ G_y(t) \\ G_z(t) \end{bmatrix}$$

is the amount of momentum created per unit time within the control volume at time t . This balance corresponds to three equations, one per component:

$$\left\{ \begin{array}{l} P_x(t_2) - P_x(t_1) - \int_{t_1}^{t_2} F_x(t) dt = \int_{t_1}^{t_2} G_x(t) dt \\ P_y(t_2) - P_y(t_1) - \int_{t_1}^{t_2} F_y(t) dt = \int_{t_1}^{t_2} G_y(t) dt \\ P_z(t_2) - P_z(t_1) - \int_{t_1}^{t_2} F_z(t) dt = \int_{t_1}^{t_2} G_z(t) dt \end{array} \right. \quad (5.5)$$

Differential form of balance laws

By considering a very short time lapse Δt , with reasoning analogous to the case of a conservation law (5.2), we can write a balance law in a different form:

Balance law in differential form

A quantity is said to satisfy a **balance law** if the following equality holds *always*:

$$\frac{dE(t)}{dt} - \Phi(t) = R(t) \quad (5.6)$$

For a vector quantity such as momentum, the differential form again corresponds to three equations, one per component:

$$\begin{cases} \frac{dP_x(t)}{dt} - F_x(t) = G_x(t) \\ \frac{dP_y(t)}{dt} - F_y(t) = G_y(t) \\ \frac{dP_z(t)}{dt} - F_z(t) = G_z(t) \end{cases} \quad (5.7)$$

Example with a moving control surface

Take a coordinate system (x, y, z) where y has a horizontal direction and z an upward direction. At a particular time instant a tennis ball has momentum $[0, 1.70, 0.98]$ N s. It flies for two seconds. During this time there's a continuous and constant supply of momentum within the ball, equal to $[0, 0, -0.579]$ N. No net momentum is leaving or entering the ball in any other way (there is a flux of momentum corresponding to air pressure, but it is such that its total over the whole surface is approximately zero).



We assume that momentum satisfies a balance law. Choose an imaginary, closed control surface that tightly wraps the tennis ball and moves with it. From the description above we have

$$t_1 = 0 \text{ s}, \quad t_2 = 2 \text{ s},$$

$$\mathbf{P}(t_1) = \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{ N s}, \quad \mathbf{F}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N (const.)}, \quad \mathbf{G}(t) = \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N (const.)}.$$

The balance law (5.4) allows us to find the amount of momentum in the tennis ball at time $t_2 = 2$ s:

$$\begin{aligned}
 \mathbf{P}(t_2) &= \mathbf{P}(t_1) + \int_{t_1}^{t_2} \mathbf{F}(t) dt + \int_{t_1}^{t_2} \mathbf{G}(t) dt \\
 &= \mathbf{P}(t_1) + \mathbf{F} \cdot (t_2 - t_1) + \mathbf{G} \cdot (t_2 - t_1) \quad (\text{because } \mathbf{F}, \mathbf{G} \text{ are constant}) \\
 &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{ Ns} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N} \cdot 2 \text{ s} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N} \cdot 2 \text{ s} \\
 &= \begin{bmatrix} 0 \\ 1.70 \\ -0.18 \end{bmatrix} \text{ Ns}
 \end{aligned}$$

We see that two seconds later the ball's momentum has the same x - and y -component, but its z -component points downward – which means that the ball is moving downward.

The description above actually has enough information for applying the balance law in differential form, formula (5.7), because the value of the influx \mathbf{F} and supply \mathbf{G} are given at every time t (they are constant). We find that, at any time t between t_1 and t_2 , the rate of change of momentum is

$$\begin{aligned}
 \frac{d\mathbf{P}(t)}{dt} &= \mathbf{F}(t) + \mathbf{G}(t) \\
 &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N} \\
 &= \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N}
 \end{aligned}$$

So we could find the momentum at time t_2 , or at any other time between t_1 and t_2 , by an easy integration:

$$\begin{aligned}
 \mathbf{P}(t_2) &= \mathbf{P}(t_1) + \int_{t_1}^{t_2} \frac{d\mathbf{P}(t)}{dt} dt \\
 &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{Ns} + \int_{t_1}^{t_2} \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{N} dt \\
 &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{Ns} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{N} \cdot (t_2 - t_1) \quad (\text{because the integrand is constant}) \\
 &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{Ns} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{N} \cdot 2 \text{s} \\
 &= \begin{bmatrix} 0 \\ 1.70 \\ -0.18 \end{bmatrix} \text{Ns}
 \end{aligned}$$

Example with a static control surface

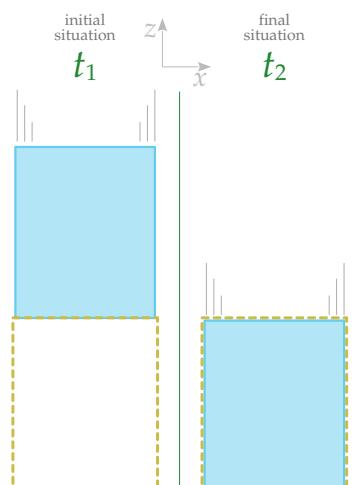
Consider again the example from p. 5.2 of a cylindrical block of ice moving downward, in a vacuum, during a lapse of time of 0.5 s, depicted in the side picture. The block of ice has a fully downward momentum of magnitude 4.72 N s at the final time.

We are now interested in the net amount of *momentum* that crosses, in a downward direction, the circular surface separating the initial and final location.

We could use the same strategy adopted in that example: considering a static control surface at the final position of the block of ice. From the description, our data are

$$t_1 = 0 \text{ s}, \quad t_2 = 0.5 \text{ s}, \quad \mathbf{P}(t_1) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{Ns}, \quad \mathbf{P}(t_2) = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{Ns},$$

But we don't have enough data to apply the balance law (5.4) for finding the integrated influx. It is also necessary to know the supply of momentum $\mathbf{G}(t)$, or at least its integral. This is an example of what was said in § 5.3: balance laws in general require more information than conservation laws.



Suppose we are told that the net amount of momentum generated *within the control volume*, during the 0.5 s, has magnitude 1.57 N s and fully downward direction. That is,

$$\int_{t_1}^{t_2} \mathbf{G}(t) dt = \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \text{Ns}$$

Now we can use the balance law (5.4) to find the integrated influx:

$$\begin{aligned} \int_{t_1}^{t_2} \mathbf{F}(t) dt &= \underline{\mathbf{P}}(t_2) - \underline{\mathbf{P}}(t_1) - \int_{t_1}^{t_2} \mathbf{G}(t) dt \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{Ns} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{Ns} - \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \text{Ns} \\ &= \begin{bmatrix} 0 \\ 0 \\ -3.15 \end{bmatrix} \text{Ns} \end{aligned}$$

Let us finish this example with a quick analysis using a *moving control surface*, wrapping the block of ice. We are told that the block initially has zero momentum. Moreover, there is no flux of momentum through a surface that moves along with the block. Our data are therefore

$$t_1 = 0 \text{ s}, \quad t_2 = 0.5 \text{ s},$$

$$\underline{\mathbf{P}}(t_1) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{Ns}, \quad \underline{\mathbf{P}}(t_2) = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{Ns}, \quad \underline{\mathbf{F}}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{N (const.)}.$$

Note that the volume integrals and fluxes are now represented by different symbols (we use an underline, but we could use any other graphical symbol, or simply change the letters themselves). This is because **they refer to different control surfaces and volumes**. We expect their values to be different from the previous ones, which referred to static surfaces. Indeed we already notice that

$$\int_{t_1}^{t_2} \mathbf{F}(t) dt = \begin{bmatrix} 0 \\ 0 \\ -3.15 \end{bmatrix} \text{Ns} \neq \int_{t_1}^{t_2} \underline{\mathbf{F}}(t) dt = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{Ns}.$$

We have enough data to find the total amount of momentum generated within the moving control volume:

$$\begin{aligned}\int_{t_1}^{t_2} \underline{\mathbf{G}}(t) dt &= \underline{\mathbf{P}}(t_2) - \underline{\mathbf{P}}(t_1) - \int_{t_1}^{t_2} \underline{\mathbf{F}}(t) dt \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ N s} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N s} - \int_{t_1}^{t_2} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N dt} \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ N s}\end{aligned}$$

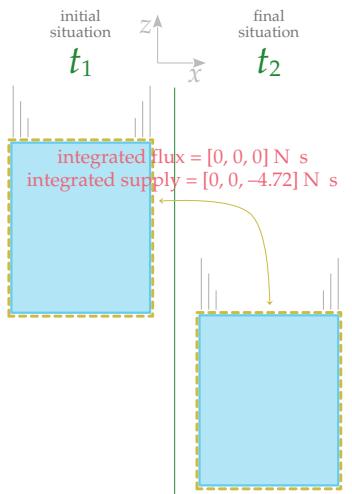
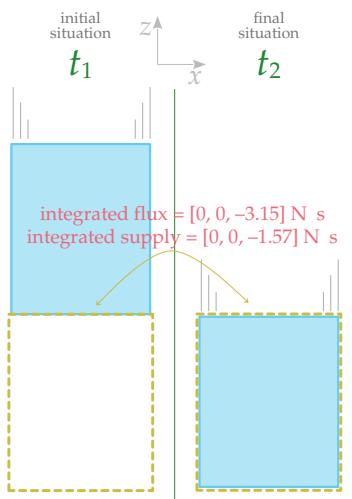
We notice again the difference:

$$\int_{t_1}^{t_2} \underline{\mathbf{G}}(t) dt = \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \text{ N s} \neq \int_{t_1}^{t_2} \underline{\mathbf{G}}(t) dt = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ N s}$$

Fluxes and supplies depend on the particular control surface and its motion

The integrated flux $\int_{t_1}^{t_2} \underline{\mathbf{F}}(t) dt$ and integrated supply $\int_{t_1}^{t_2} \underline{\mathbf{G}}(t) dt$ that appear in conservation and balance laws are strictly dependent on the sequence of closed control surfaces and volumes that we choose.

Therefore, if we analyse the same physical phenomenon with a new set of control surfaces, in general we cannot expect the results of calculations with the old set to be valid for the new one.



Exercise 5.2

Consider once more the block of ice analysed in the previous example.

This time choose a *static* closed control surface that coincides with the *initial* position of the block. This surface therefore includes all ice at time t_1 , but is empty – and has zero momentum – at time t_2 .

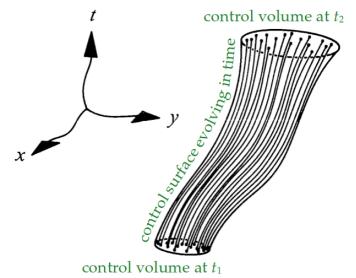
Calculate the integrated supply for this new control surface between t_1 and t_2 .

(Hint: consider the result from the exercise and the **symmetry of fluxes** for the flux through the bottom part of the surface, as well as the fact that the fluxes through the side and top surfaces are zero.)

Balance laws in General Relativity

Conservation and balance laws appear even simpler from the point of view of Relativity Theory. From a four-dimensional spacetime perspective, a 3D volume is a region, called *hypersurface*, having one less dimension than spacetime. But a sequence of 2D surfaces through time is also just a region having one less dimension than spacetime – two spatial dimensions and one temporal one. Thus the distinction between the 3D region at t_1 , the sequence of 2D surfaces between t_1 and t_2 , and the 3D region at t_2 disappear: they are seen to be just different parts of the same three-dimensional hypersurface. We perceive some parts of this hypersurface as belonging “to the same time”, showing their three dimensions all at once; and other parts as extending through time, showing only two dimensions at any time. In fact, different observers make this division in different ways.

And from a spacetime perspective, the amount of a quantity $E(t_1)$ or $E(t_2)$ within a 3D region is seen as a flux through time; so its apparent difference from the flux Φ also disappear.



Spacetime representation of the evolution of a closed control surface, containing some pointlike objects (adapted from Misner et al. 2017)

What about the magnetic flux?

In the case of magnetic flux, the idea of a conservation law is analogous, but is formulated with one less spatial dimension: we consider a closed 1D curve at t_1 , one at t_2 , and a sequence in between these times. The magnetic flux turns out to be a quantity for which it's possible to ask how much of it is “linked” to a closed curve, and how much of it is crossing a closed curve. One way to understand this is to imagine magnetic flux as a bundle of tubes or lines that are either closed or extend to infinity. It is a very fascinating quantity, and one may wonder if other quantities exist which satisfy similar balances even with one less dimension. But we shall not pursue magnetic flux and related topics in these notes.

5.4 Constitutive relations

In the previous sections we have seen several examples of use of conservation and balance laws. They allowed us to determine or predict the amount of quantities in given spatial regions or crossing through surfaces.

The examples also showed that extra information is always needed in order to use the balance laws. For instance, in the example with the tennis

ball of p. 85 we said that there was no flux of momentum through the moving control surface (the flying ball). In the example with the block of ice of p. 87 we said “where there’s no matter, there’s no momentum”; and we also said that the flux of momentum was zero on some surfaces, but non-zero on others.

Well, how did we know?

This kind of information and physical properties were *not* something we could deduce from the conservation or balance laws.

Indeed those were examples of physical properties that apply, not universally, but only to some specific phenomena, and only to some theory-dependent degree of approximation. They were examples of *constitutive relations*:

Constitutive relation

A **constitutive relation**, also called *constitutive equation* or *closure equation*, is a physical relationship or property that is only true for specific physical phenomena.

A constitutive relation often has different expressions, of varying levels of approximation, in different theories.

Constitutive relations express the diversity that we observe around us, for example the fact that a body of water can easily change shape upon movement, as opposed to a block of concrete, which we rely on for its rigidity. The differences between states of matter – solid, liquid, gas, plasma, and there are others – arise from different constitutive relations. Constitutive relations also mark the difference between specialized or approximate physical theories, for example between Newtonian mechanics, which applies for low speeds and weak gravitational phenomena only, and general relativity, which applies also to cosmological scales.

When we read that a new physical phenomenon has been discovered, usually that means that a new *constitutive relation* has been discovered. Depending on the specific scientific field you’ll work in, you’ll learn some constitutive relations in more detail than others.

Constitutive relations come in a great variety of mathematical forms. Some of them are simple algebraic relations between the volume integral of one quantity and the flux of another. Others involve spatial or time derivatives. Other still involve integrals in space or in time.

“Having the power of constituting, establishing, or giving formal, definite, or organized existence to something; constructive.”

“That makes a thing what it is; forming an essential part or element; essential.”

(‘Constitutive’, Oxford English Dictionary)



Four states of matter (image by Spirit469⁴)

Examples

Let us briefly discuss some of the constitutive equations that have silently been used in the examples of the previous sections. All the constitutive relations below are valid only in “Newtonian approximation”, that is, when speeds are much lower than the speed of light and gravitation is weak.

Constitutive relation for mass-energy and matter. If a small control volume contains an amount of matter N , then it also contains an amount of mass-energy

$$M = \rho N$$

where ρ , called *molar mass*, is (approximately) a constant that depends on the kind of matter. This constitutive relation is the reason why an amount of matter is often quantified in terms of mass.

Constitutive relation for momentum and matter. Take a small control volume containing an amount of matter N , and such that there is no flux of matter across its surface. Then this control volume also contains an amount of momentum

$$\mathbf{P} = M\mathbf{v}$$

This is the famous relation that old textbooks take as the *definition* of momentum.

where \mathbf{v} is the velocity of the control volume. Note that this relation uses the previous one between mass and matter, so we have $\mathbf{P} = \rho N\mathbf{v}$.

More symmetric relation between mass & momentum and matter

When we work with *densities*, which express the amounts of a quantity per unit volume or per unit surface, the relation between mass-energy & momentum on one side, and matter & its flux on the other, becomes more explicit and symmetric. We have

$$m = \rho n \quad \mathbf{p} = \rho \mathbf{j}$$

where m and \mathbf{p} are the mass and momentum densities, and n and \mathbf{j} are the matter and matter-flux densities.

Constitutive relation for supply of momentum. If a small control volume contains an amount of mass-energy M , then it also has a supply of momentum

$$\mathbf{G} = Mg$$

where \mathbf{g} is a vector that depends on the gravitational field. This constitutive relation expresses the *gravitational force*.

For physical phenomena on the Earth's surface, the vector \mathbf{g} is approximately constant: it points towards the Earth's centre and has magnitude $g := |\mathbf{g}| \approx 9.8 \text{ N/kg} \equiv 9.8 \text{ m/s}^2$.

For physical phenomena further away from Earth, for instance when dealing with satellite motion, the vector \mathbf{g} points towards the Earth's centre and has magnitude

$$g = \frac{GM_E}{r^2}$$

where $G = 6.67 \text{ N m}^2/\text{kg}^2$ is the gravitational constant, $M_E = 5.97 \times 10^{24} \text{ kg}$ is [Earth's mass](#)⁶, and r is the distance of the control volume from the Earth's centre.



(image from NASA⁵)

Constitutive relation for energy-mass of matter. Take a small control volume on Earth's surface containing an amount of matter N , and such that there is no flux of matter across its surface. Then this control volume also contains an amount of energy-mass

$$E = E_0 + U + \frac{1}{2}Mv^2 + Mgh$$

where the first term is a constant zero level of energy, the second term U is called *internal energy*, the third is called *kinetic energy*, and the fourth *gravitational potential energy*. The internal energy U is typically a function of the amount of matter N .

Matter enters this formula also through the constitutive relation $M = \rho N$. The zero value E_0 is usually unimportant because we mostly work with energy changes. This value is huge in ordinary situations, as discussed in § 3.4.

5.5 The Seven Wonders of the World

In ch. 3 we discussed seven primitive quantities that are common to all our main physical theories. The reason for introducing those seven primitive quantities is that each of them obeys a balance law. More precisely: *matter*, *electric charge*, and *magnetic flux* actually obey strict conservation laws, like (5.1). Whereas *energy*, *momentum*, *angular momentum*, and *entropy*

satisfy balance laws, like (5.3), that may have a supply; only in special circumstances these quantities also obey conservation laws.

What's remarkable is that these seven balance laws are known, so far, to be satisfied by *all* phenomena and in *all* our main physical theories, approximate or not: from subatomic scales to cosmological scales; from Newtonian mechanics to general relativity and quantum theory. No exceptions are known. They are our fundamental laws:

The seven fundamental balance laws

- Conservation of matter**
- Conservation of electric charge**
- Conservation of magnetic flux**
- Balance of energy**
- Balance of momentum**
- Balance of angular momentum**
- Balance of entropy**

These balances are truly the Seven Wonders of the World, even more long-lasting than [the traditional “seven wonders”⁷](#).

Once you learn these balances, you can apply them to any kind of physical phenomenon or problem you work with: construction of bridges, control of chemical reactions, operation of GPS navigation and satellites, monitoring of nuclear power plants, sending robots to Mars, or collisions of subatomic particles.

Roles of the seven fundamental balances in description and prediction

The seven fundamental balances govern every physical phenomenon. Yet this doesn't mean that all of them are always used explicitly.

For some physical phenomena all the seven laws enter our calculations.

For some other physical phenomena, some of the seven laws do not appear explicitly in our calculations; yet they still enter implicitly in the way we choose to set up or describe the phenomenon. For instance we may choose control volumes or control surfaces in such a way that some conservation laws are automatically satisfied. A typical example is the choice of control surface around a given object (amount of matter), which guarantees that the law of conservation of matter is automatically satisfied.

For still other physical phenomena, some of the seven laws may not be required because we do not need the kind of physical information they



The description of how a common lighter works, thanks to piezoelectricity⁸, requires more or less all seven fundamental laws to be explicitly accounted for.

provide. We saw an example of this with the flat-tyre problem of p. 79, where we only used the conservation of matter but we weren't interested in what happened to other quantities such as energy or momentum.

Some physical phenomena may also be predicted using either a particular set of fundamental laws, for instance conservation of matter and balance of momentum; or a different set, for instance conservation of matter and balance of energy. We shall see examples of all these possibilities in the next sections.

The fact that the seven fundamental balances govern every physical phenomenon doesn't mean that they can be used alone, by themselves. We saw examples of this fact in ch. 5, where the fundamental laws often needed to be augmented by appropriate constitutive equations.

In the following chapters we shall explore the seven fundamental balance laws in more detail, also giving examples of constitutive relations that typically accompany them.

URLs for chapter 5

1. <https://pglpm.github.io/7wonders/media/surfacemotion.mp4>
2. <https://pubchem.ncbi.nlm.nih.gov/element/Argon>
3. <https://encyclopedia.che.ingen.umich.edu/distillation-columns/>
4. https://commons.wikimedia.org/wiki/File:Four_Fundamental_States_of_Matter.png
5. https://nssdc.gsfc.nasa.gov/imgcat/html/group_page/ER.html
6. <https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>
7. <https://education.nationalgeographic.org/resource/seven-wonders-ancient-world/>
8. <https://www.britannica.com/science/piezoelectricity>

Chapter 6

Conservation of matter

6.1 Formulation and generalities

Choose an arbitrary sequence of closed control surfaces between coordinate times t_1 and t_2 , enclosing a sequence of control volumes. Denote by $N(t)$ and $J(t)$ the amount of matter in the control volume and the influx through the control surface at time t . Then the law of **conservation of matter** is expressed by either of the equations

Conservation of matter

$$N(t_2) - N(t_1) - \int_{t_1}^{t_2} J(t) dt = 0 \quad \text{integral form} \quad \frac{dN(t)}{dt} - J(t) = 0 \quad \text{differential form} \quad (6.1)$$

Conservation of matter is a law that we intuitively take for granted and use continuously in our life. The very notion of ‘object’ – including living objects – is possible thanks to this fundamental regularity of nature: we can speak of objects because they exist for some time and we can follow them as they move in space. This is possible because the integrated flux through a small surface of the “stuff” that makes the object equals the amount that disappeared in a nearby small region: we can then say that the “stuff” has moved.

In fact even the notion of *velocity* of an object is actually built from the notions of flux and of volume integral, but it would make little sense if the flux didn't equal the rate of change of the volume integral.

Velocity from flux and volume integral

Suppose we have a coordinate system (t, x, y, z) . Consider a small spatial region of cuboid shape delimited by two surfaces having x constant, two having y constant, and two having z constant. The instantaneous velocity component v_x is defined as the ratio of: (a) the flux J_x through any of the two x -constant surfaces, crossing in the positive- x orientation, divided by surface area A , and (b) the volume integral N in the cuboid region, divided by the region's volume V :

$$v_x := \frac{J_x/A}{N/V}$$

And analogously for the other two velocity components. This relation is valid as a first approximation for enough small velocities and weak gravitational fields (Newtonian approximation).

6.2 Examples of constitutive relations

Relation between matter and mass-energy

The most common constitutive relation used with conservation of matter is the one relating a constant amount of mass-energy M to a particular amount of matter N :

Molar mass

$$M = \rho N \quad (6.2)$$

where the proportionality constant $\rho = M/N$ is called **molar-mass constant**, and depends on the kind of matter considered.

For instance, in many physical phenomena involving water we assume the constitutive relation above, with a of [water molar-mass constant¹](#) of approximately $\rho_{\text{H}_2\text{O}} = 0.0180 \text{ kg/mol}$. If a volume contains $N = 20 \text{ mol}$ of water, we usually attribute to it a mass-energy of

$$\begin{aligned} M &= \rho_{\text{H}_2\text{O}} N \\ &= 0.0180 \text{ kg/mol} \times 20 \text{ mol} \\ &= 0.36 \text{ kg}. \end{aligned}$$

The exact value of the molar-mass constant depends not only on the substance but also on the context and application: the substance may

actually consist of a mixture of different kinds of matter. ‘Air’ for example is a mixture of different chemical elements, and their proportion in the mixture may depend on physical conditions such as temperature, and on geographical position. ‘Water’ is a mixture of different isotopes², and again the mixture proportions may vary with physical conditions.

The constant relating mass-energy and amount of matter is the same for volume integrals and for fluxes. So to a matter flux J we can associate a mass-energy flux ρJ . For instance, if through a surface there’s a flux of $J = 30 \text{ mol/s}$ of water, then we associate to it a mass flux of

$$\rho_{\text{H}_2\text{O}} J = 0.0180 \text{ kg/mol} \times 30 \text{ mol/s} = 0.54 \text{ kg/s}.$$

Using this constitutive relation we can re-express the conservation of matter in terms of mass. Suppose that for a given kind of matter the proportionality constant is ρ ; then conservation of matter implies

$$\rho N(t_2) - \rho N(t_1) - \int_{t_1}^{t_2} \rho J(t) dt = 0 \quad \frac{d\rho N(t)}{dt} - \rho J(t) = 0$$

This is the ‘law of conservation of mass’. Keep in mind that this is not a universal law: it’s only an approximate law, only valid as long as the constitutive relation between matter and mass-energy is acceptable.

Phenomena for which this constitutive equation is no longer valid appear for example in particle physics, nuclear physics, and astrophysics.

Exercise 6.1

At a particular time, a party balloon contains 0.012 kg of Helium. A minute later, the same balloon contains 0.010 kg of helium. Assume conservation of Helium, as well as a constitutive relation between amount and mass with [Helium molar-mass constant³](#) $\rho_{\text{He}} = 4 \times 10^{-3} \text{ kg/mol}$.

1. How much is the *integrated efflux* of Helium, **in moles**, through the balloon’s surface during the one-minute lapse of time?
2. Assume that the efflux of Helium was constant in time during the one-minute time lapse. How much was the efflux of Helium, in moles/second?



Relation between matter and momentum

Closely related to the previous constitutive relation is another one relating a amount of matter and an amount of momentum. Assume that the constitutive relation between matter and mass, previously discussed, is valid, and associates a mass-energy $M = \rho N$ to the amount of matter N .

Newtonian formula for momentum

Consider a small control volume containing, at time t , an amount N of matter of a particular kind. If the volume is small enough we can consider all matter within as moving with approximately the same velocity \mathbf{v} . The **Newtonian formula for the momentum of matter** is a constitutive relation that says that the control volume also contains an amount of momentum

$$\mathbf{P} = M\mathbf{v} \equiv \rho N\mathbf{v} \quad (6.3)$$

(The velocity \mathbf{v} is actually determined by the flux of matter, see the blue box on p. 98, but for the present purpose you can use your intuitive understanding of this velocity).

An analogous constitutive relation also exists between flux of momentum and flux of matter. Its most general formulation involves many mathematical terms, so let's consider a simplified situation instead.

Convective flux of momentum (special case)

Take a small surface that is *not moving* in the coordinate system (t, x, y, z) , and is parallel to the coordinates y and z . Call J_x the flux of matter through this surface in the positive- x direction. Then through this surface there's also a flux of momentum

$$\mathbf{F}_c = \rho J_x \mathbf{v} \quad (6.4)$$

called a **convective flux of momentum**. Analogous relations hold for small, non-moving surfaces parallel to the zx and xy planes.



Keep in mind that usually there are additional, non-convective flows of momentum (forces) associated with matter, as we shall see later.

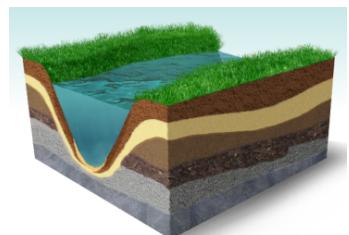
These constitutive relations are only approximate and become incorrect when the speeds involved are close to the speed of light or the gravitational fields are strong, or when some kind of electromagnetic phenomena are also involved.

Exercise 6.2

The side picture shows the section of a river. The river has an approximately *steady flow* of water; this means that if we choose a static control volume anywhere in the river, such as the one in the depicted section, the amount of water N therein is *constant in time*.

Assume that (a) water satisfies conservation of matter; (b) the amount of water in the illustrated volume is $N = 4 \times 10^5 \text{ mol}$; (c) water only flows across the imaginary vertical surfaces that delimit the river section on the front and back of the picture, cutting across the river.

1. What can you say about the two fluxes across the two imaginary vertical surfaces, each considered as crossed from right to left in the picture? Can you find their numerical values?
2. You are told that the illustrated volume of water has total momentum $P = 21\,600 \text{ N s}$, directed from left to right perpendicularly to the imaginary vertical surfaces. Assuming the Newtonian formula for momentum, and the molar-mass constant $\rho_{\text{H}_2\text{O}} = 0.0180 \text{ kg/mol}$, can you find the values of the fluxes from the previous question?



(Illustration by Cathryn Briggs⁴)

Thermodynamics

Matter enters in many constitutive relations in thermodynamics and chemistry. An example is the ideal gas law

$$pV = RNT$$

relating the pressure p (essentially a momentum flux), volume V , temperature T of an amount N of an ideal gas in equilibrium conditions; $R = 8.314\,462\,618\,153\,24 \text{ N m/(K mol)}$ (exactly) is the **molar gas constant⁵**.

6.3 Examples of applications

Rigid-body and particle mechanics

In many applications, we set up our description of the physical phenomenon of interest in such a way that the law of conservation of matter is automatically satisfied. The law thus enter our descriptions and predictions silently, in an implicit way.

One example is when we focus on a small (relative to the environment) object, such as a tennis ball in an open field, or a planet in a solar system.

*** In these cases we can choose a sequence of control surfaces

there is an amount of matter that at each time occupies a small volume (relative to the environment). We therefore decide to use a sequence of closed control surfaces (§ 4.1) that tightly encloses the amount of matter at each time t . As the amount of matter in these control volumes is constant, that is, $dN/dt = 0$, by the law of conservation of matter the flux must then be zero: $J(t) = 0$. The requirement that the flux be zero then determines how the closed control surfaces must move in space; this we finally encode in the velocity of the small object. Once this is done, the law of conservation of matter is automatically taken care of, and we can focus on other conservation laws.

The same procedure is used in the description of a rigid body, that is, an amount of matter that occupies a region of constant volume and shape. The volume doesn't need to be small in this case. Again by using tightly enclosing control surfaces, their motion is determined by

In chemistry

URLs for chapter 6

1. <https://webbook.nist.gov/cgi/inchi/InChI%3D1S/H20/h1H2>
2. <https://www.britannica.com/science/isotope>
3. <https://webbook.nist.gov/cgi/inchi/InChI%3D1S/He>
4. <https://dribbble.com/squarefountain>
5. <https://www.britannica.com/science/universal-gas-constant>

Chapter 7

Balance of energy

Chapter 8

Balance of momentum

Chapter 9

Balance of angular momentum

Chapter 10

Densities and flux-densities

How do we mathematically represent the seven main quantities? We said that, for each, we can ask *how much in this region?* and *how much through this surface during this time?* We must therefore use mathematical objects that allow us to answer these two questions. These objects are the *density* and the *flux vector*. The idea behind them is quite simple.

Density

Take a point in spacetime with coordinates (t_0, x_0, y_0, z_0) . At the instant t_0 , imagine a very small cuboid centred at (x_0, y_0, z_0) , with sides aligned with the coordinate axes x, y, z and of lengths $\Delta x, \Delta y, \Delta z$. Then we have

$$\text{amount of quantity in cuboid} = n(t_0, x_0, y_0, z_0) \Delta x \Delta y \Delta z . \quad (10.1)$$

where $n(t_0, x_0, y_0, z_0)$ is the **density** at the spacetime point (t_0, x_0, y_0, z_0) .

The density tells us how much quantity there's in a unit of volume. As the notation suggests, it is a function of the coordinates, that is, of time and spatial position. This functional dependence reflects the fact that we can have a larger amount of a quantity concentrated in some regions at some times, than in other regions or at other times.

In order to calculate the total amount of the quantity in an arbitrary 3D region, we simply divide it into very small cuboids similar to the one above. For each cuboid we calculate the respective amount; this amount will generally be different from cuboid to cuboid, because the value of the

density $n(t, x, y, z)$ will be different at each cuboid's centre. Then we sum up all these amounts.

Isn't this an integral?

You probably recognize this procedure as the description of integration. Indeed we can write:

$$\text{total amount of quantity in 3D region} = \iiint_{\text{3D region}} n(t, x, y, z) dx dy dz . \quad (10.2)$$

Clearly we must know the value of the density n at all points within the region, in order to calculate the total amount. Note that if the quantity is absent in some subregion, then $n(\dots) = 0$ there.

Do not confuse flux with movement

Flux

For the flux of the quantity through a surface we consider three different cases:

x-Flux

Take again a point in spacetime with coordinates (t_0, x_0, y_0, z_0) . Keeping x_0 fixed, imagine a very small rectangular surface centred at (x_0, y_0, z_0) , with sides aligned with the coordinate axes y, z and of lengths $\Delta y, \Delta z$. Imagine that this surface exists for a lapse of time Δt around the time t_0 . Then we have

$$\left. \begin{array}{l} \text{amount of quantity} \\ \text{through rectangle } \Delta y \Delta z \\ \text{towards positive } x \\ \text{during time lapse} \end{array} \right\} = j_x(t_0, x_0, y_0, z_0) \Delta t \Delta y \Delta z . \quad (10.3)$$

where $j_x(t_0, x_0, y_0, z_0)$ is the **x-flux** at the spacetime point (t_0, x_0, y_0, z_0) .

The **x-flux** tells us how much quantity is flowing in a unit of time through a unit of surface parallel to y, z . Also the **x-flux** is a function of the coordinates, because the flux could be larger through some surfaces at some times, than through other surfaces at other times. If the quantity is

flowing in the negative x direction, then j_x will be negative; and clearly if no quantity is flowing through the small rectangle, then $j_x = 0$.

In an analogous way we define a y -flux and a z -flux:

y -Flux

$$\left. \begin{array}{l} \text{amount of quantity} \\ \text{through rectangle } \Delta z \Delta y \\ \text{towards positive } y \\ \text{during time lapse} \end{array} \right\} = j_y(t_0, x_0, y_0, z_0) \Delta t \Delta z \Delta y . \quad (10.4)$$

z -Flux

$$\left. \begin{array}{l} \text{amount of quantity} \\ \text{through rectangle } \Delta x \Delta y \\ \text{towards positive } z \\ \text{during time lapse} \end{array} \right\} = j_z(t_0, x_0, y_0, z_0) \Delta t \Delta x \Delta y . \quad (10.5)$$

The three fluxes can be grouped into the **flux vector**

$$\mathbf{j}(t, x, y, z) := (j_x(t, x, y, z), j_y(t, x, y, z), j_z(t, x, y, z)) \quad (10.6)$$

In order to calculate the total amount of the quantity through an arbitrary sequence of 2D surfaces during a time lapse, we simply divide the total time into very brief time lapses, and for each of these we approximate the surface with very small rectangles aligned along the three axes. For each such lapse and rectangle we calculate the respective flux amount, using j_x , j_y , or j_z depending on the orientation of the rectangle. Then we sum up all these amounts.

T

This is also the description of an integration, which can be written

total amount of quantity through sequence of surfaces =

$$\begin{aligned} & \iiint j_x(t, x, y, z) dt dy dz + \\ & \iiint j_y(t, x, y, z) dt dz dy + \\ & \iiint j_z(t, x, y, z) dt dx dy \end{aligned} \quad (10.7)$$

We shall not have to calculate density integrals (10.2) and flux integrals (10.7) in these notes. In concrete physics problems these integrals are often difficult to calculate, and require advanced mathematical techniques and specialized software. But if you'll ever end up working in fields such as computational fluid dynamics, atmospheric or ocean modelling, or numerical relativity, then you shall probably encounter them again.

The definitions above of density and flux are appropriate if the quantity in question is a scalar. For vector quantities such as momentum and angular momentum, the density is a vector, and each of the three fluxes is a vector as well.

10.1 Balance laws expressed with derivatives

In ch. 10 we saw that each of the main seven quantities is represented by a density and by a flux vector, which allow us to calculate the amount of quantity within a 3D region at a particular time, and flowing through a 2D surface during an time interval. These amounts are exactly what $N(t_1)$, $N(t_2)$, J express. Therefore, a conservation or balance law must also give some relationship between density and flux vector.

The way to find such relationship is intuitive and mathematically simple, if only a bit lengthy. The idea is to formulate the balance law for a very simple case, the same simple set-up that we used in ch. 10 to introduce density and flux.

Take a point in spacetime with coordinates (t, x, y, z) . We apply the conservation law (5.1) to the following situation:

$N(t_1)$: Choose an initial time $t_1 = t - \frac{\Delta t}{2}$, slightly before t , and a cuboid region centred at (x, y, z) with sides $\Delta x, \Delta y, \Delta z$. The total amount of the quantity within this region is given, adapting formula (10.1), by

$$N(t_1) = n\left(t - \frac{\Delta t}{2}, x, y, z\right) \Delta x \Delta y \Delta z. \quad (10.8)$$

$N(t_2)$: Choose a final time $t_2 = t + \frac{\Delta t}{2}$, slightly after t , and a cuboid region centred at (x, y, z) with sides $\Delta x, \Delta y, \Delta z$. The total amount of the quantity within this region is given, adapting formula (10.1), by

$$N(t_2) = n\left(t + \frac{\Delta t}{2}, x, y, z\right) \Delta x \Delta y \Delta z. \quad (10.9)$$

J : We calculate the flux separately through the six rectangular surfaces bounding the small cuboid region:

J_x : First choose a rectangular surface centred slightly to one side of the point (x, y, z) : at $(x - \frac{\Delta x}{2}, y, z)$, with sides parallel to y, z and lengths $\Delta y, \Delta z$. We keep this surface constant for the small duration Δt . The x -flux through this rectangle, towards positive x , during Δt , is according to formula (10.3)

$$J_x^{\text{in}} = j_x(t, x - \frac{\Delta x}{2}, y, z) \Delta t \Delta y \Delta z .$$

Now choose the rectangular surface parallel to the previous one, but on the other side of the point (x, y, z) , centred at $(x + \frac{\Delta x}{2}, y, z)$. The x -flux through this rectangle, towards positive x , during Δt , is

$$J_x^{\text{out}} = j_x(t, x + \frac{\Delta x}{2}, y, z) \Delta t \Delta y \Delta z .$$

Note that the flux J_x^{in} points to the interior of the 3D region, towards its centre (x, y, z) ; whereas J_x^{out} points to the exterior, away from the centre. The total flux into the region is therefore given by their difference:

$$\begin{aligned} J_x &= J_x^{\text{in}} - J_x^{\text{out}} \\ &= [j_x(t, x - \frac{\Delta x}{2}, y, z) - j_x(t, x + \frac{\Delta x}{2}, y, z)] \Delta t \Delta y \Delta z \end{aligned} \quad (10.10)$$

J_y, J_z : An analogous reasoning can be made to find the total flux through the other four surfaces, two parallel to z, y and two to x, y . We find

$$\begin{aligned} J_y &= J_y^{\text{in}} - J_y^{\text{out}} \\ &= [j_y(t, x, y - \frac{\Delta y}{2}, z) - j_y(t, x, y + \frac{\Delta y}{2}, z)] \Delta t \Delta z \Delta x \end{aligned} \quad (10.11)$$

$$\begin{aligned} J_z &= J_z^{\text{in}} - J_z^{\text{out}} \\ &= [j_z(t, x, y, z - \frac{\Delta z}{2}) - j_z(t, x, y, z + \frac{\Delta z}{2})] \Delta t \Delta x \Delta y \end{aligned} \quad (10.12)$$

The total flux into the region during the time interval Δt is finally the sum of the three above:

$$J = J_x + J_y + J_z \quad (10.13)$$

Up to now we have only chosen arbitrary surfaces, regions, and a time interval. Now let's assume that the corresponding amounts and fluxes

satisfy a balance law like (5.1). Putting together the puzzle pieces we find

$$N(t_2) - N(t_1) - J = 0 \quad \implies$$

$$\left. \begin{aligned} & n\left(t + \frac{\Delta t}{2}, x, y, z\right) \Delta x \Delta y \Delta z - n\left(t - \frac{\Delta t}{2}, x, y, z\right) \Delta x \Delta y \Delta z \\ & - [j_x(t, x - \frac{\Delta x}{2}, y, z) - j_x(t, x + \frac{\Delta x}{2}, y, z)] \Delta t \Delta y \Delta z \\ & - [j_y(t, x, y - \frac{\Delta y}{2}, z) - j_y(t, x, y + \frac{\Delta y}{2}, z)] \Delta t \Delta z \Delta x \\ & - [j_z(t, x, y, z - \frac{\Delta z}{2}) - j_z(t, x, y, z + \frac{\Delta z}{2})] \Delta t \Delta x \Delta y \end{aligned} \right\} = 0$$

This expression is quite long, but it should be intuitively understandable if you try to identify the individual summed terms.

Now we take the last, long expression – which is an equality – and divide its left and right sides by $\Delta t \Delta x \Delta y \Delta z$. Note that many of the “ Δ ” factors simplify out. We arrive at this:

$$\left. \begin{aligned} & \frac{n\left(t + \frac{\Delta t}{2}, x, y, z\right) - n\left(t - \frac{\Delta t}{2}, x, y, z\right)}{\Delta t} \\ & + \frac{j_x\left(t, x + \frac{\Delta x}{2}, y, z\right) - j_x\left(t, x - \frac{\Delta x}{2}, y, z\right)}{\Delta x} \\ & + \frac{j_y\left(t, x, y + \frac{\Delta y}{2}, z\right) - j_y\left(t, x, y - \frac{\Delta y}{2}, z\right)}{\Delta y} \\ & + \frac{j_z\left(t, x, y, z + \frac{\Delta z}{2}\right) - j_z\left(t, x, y, z - \frac{\Delta z}{2}\right)}{\Delta z} \end{aligned} \right\} = 0 .$$

Still a long expression; but examine each fraction: we have a difference between a term calculated at some point, and one calculated at some point plus an increment “ Δ ”; this difference is then divided by that increment. This is the definition of *partial derivative*, that is, the derivative with respect to one variable, while keeping the other variables fixed. The short symbol for this is “ $\frac{\partial \dots}{\partial \dots}$ ”, so we can simply write

Conservation law in derivative form

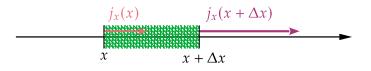
$$\frac{\partial n}{\partial t} + \frac{\partial j_x}{\partial x} + \frac{\partial j_y}{\partial y} + \frac{\partial j_z}{\partial z} = 0 \quad (10.14)$$

Let's make a simple check to see if this formula makes sense, in a simple

one-dimensional case where we disregard the y, z coordinates (let's say their fluxes are zero), so that the balance law becomes

$$\frac{\partial n}{\partial t} + \frac{\partial j_x}{\partial x} = 0.$$

Suppose that the derivative of the x -flux is positive: $\frac{\partial j_x}{\partial x} > 0$. This means that the x -flux increases as we move from x to $x + \Delta x$, as shown in the side figure. The x -flux at x is bringing some amount of quantity into the region of interest, the x -flux at $x + \Delta x$ is taking a larger amount out of that region. Therefore, during the short time Δt in which this flux takes place, the total amount of quantity within the region will decrease. But the total amount is given by $n(t)$, which is therefore decreasing with time. This means that its time derivative is negative: $\frac{\partial n}{\partial t} < 0$. This agrees with balance law above: the sum of the two derivatives must be zero, so if one is positive, the other must be negative.



Balance laws like (10.14) are the basis of many important computational and simulation methods in a huge variety of physical applications: simulation of the ocean around an offshore oil platform, of wind in a wind farm, of air around an aeroplane's wing, of earthquakes, of electromagnetic-wave propagation, of oscillations in a suspension bridge, of percolation of fluids through terrain, of chemical reactions, of energy transfer... The list could go on for pages!

The basic idea for using a balance law for a simulation is easy to understand. Take again the simpler one-dimensional formula without y, z coordinates, and rewrite it remembering the basic meaning of the derivative:

$$\frac{n(t + \Delta t, x) - n(t, x)}{\Delta t} + \frac{j_x(t, x + \frac{\Delta x}{2}) - j_x(t, x - \frac{\Delta x}{2})}{\Delta x} = 0$$

Suppose we know all quantities at time t ; we can then find the density n at time $t + \Delta t$. Leave it on the left side and bring all other terms on the right side, multiplying them by Δt :

$$n(t + \Delta t, x) = \Delta t \left[n(t, x) - \frac{j_x(t, x + \frac{\Delta x}{2}) - j_x(t, x - \frac{\Delta x}{2})}{\Delta x} \right]$$

which is essentially [Euler's method](#)¹.

This formula can then be used iteratively to find the density n at later times. As a simple simulation set-up, we consider a region of space of interest and divide it into cells of width Δx . We also treat time in steps of Δt . Then the iterative procedure goes as follows:

1. *Initial values:* Assign the value of n at each x -cell at the initial time.
2. *For-loop in x :* for each x -cell, calculate the value of n at $t + \Delta t$, using the formula above.
3. Increase time by Δt .
4. Go to 2.

For example, suppose we have two x -cells of width $\Delta x = 1$: one centred at x_1 and the other at $x_2 = x_1 + 1$. There are three boundaries: the leftmost at $x_1 - \frac{1}{2}$, one between the cells at $x_1 + \frac{1}{2} \equiv x_2 - \frac{1}{2}$, and the rightmost at $x_2 + \frac{1}{2}$. Then we have the following steps:

1. Assign the values $n(t_0, x_1), n(t_0, x_2)$ of n at x_1 and x_2 at time t_0 .
2. Assign the values

$$j_x(t_0, x_1 - \frac{1}{2}), j_x(t_0, x_1 + \frac{1}{2}), j_x(t_0, x_2 + \frac{1}{2})$$

of the fluxes at the three boundaries at time t_0 .

3. Calculate the values of n at x_1 and x_2 at time $t_1 = t_0 + \Delta t$, for instance

$$n(t_1, x_1) = \Delta t \left[n(t_0, x_1) - \frac{j_x(t_0, x_1 + \frac{1}{2}) - j_x(t_0, x_1 - \frac{1}{2})}{\Delta x} \right].$$

Note that we have all quantities required on the right side.

4. Assign the values

$$j_x(t_1, x_1 - \frac{1}{2}), j_x(t_1, x_1 + \frac{1}{2}), j_x(t_1, x_2 + \frac{1}{2})$$

of the fluxes at the three boundaries at time t_1 .

5. Calculate the values of n at x_1 and x_2 at time $t_2 = t_1 + \Delta t$, for instance

$$n(t_2, x_1) = \Delta t \left[n(t_1, x_1) - \frac{j_x(t_1, x_1 + \frac{1}{2}) - j_x(t_1, x_1 - \frac{1}{2})}{\Delta x} \right].$$

Note that we have again all quantities required on the right side.

6. And so on.

This routine requires us to know the flux at each x -cell and at each time. But we shall soon see that this requirement can be bypassed.

The rudimentary routine above is the basis from which more precise and complex simulation routines are developed.

Exercise 10.1

Write a script, in your preferred programming language, that implements the simulation routine above in an x -grid with four cells $x = 1, \dots, x = 4$. Take a grid size $\Delta x = 1$ and time step $\Delta t = 1$. We can imagine that the quantity in question is electric charge. Use the following initial values for $n(t, x)$ at $t = 0$:

$$n(t=0, x=1) = 7, \quad n(t=0, x=2) = 0, \quad n(t=0, x=3) = 0, \quad n(t=0, x=4) = 7$$

and assume the flux is, at each time t :

$$\begin{aligned} j_x(t, x=0.5) &= 0, \\ j_x(t, x=1.5) &= +2, \quad j_x(t, x=2.5) = 0, \quad j_x(t, x=3.5) = -2, \\ j_x(t, x=4.5) &= 0. \end{aligned}$$

Questions and tasks:

1. Which time evolution for n do you observe? Does it make sense, given the fluxes above?
2. Feel free to try out or generalize your script with:
 - different initial values
 - different fluxes
 - an explicit formula for the flux, for instance

$$j_x(t, x) = 8x^3 - 60x^2 + 118x - 45$$

- time-dependent fluxes
- more cells
- two or three dimensions, including y - or z -fluxes

URLs for chapter 10

1. <https://mathworld.wolfram.com/EulerForwardMethod.html>

Chapter 11

Matter

Matter is probably the easiest quantity to grasp intuitively; it's what we can really call "stuff". It is important to clearly distinguish matter from *mass*: mass can be considered a property of matter, but the two are different: for example, the mass of some amount of matter may change, even if the amount of matter stays the same.

There are different kinds of matter, each of which satisfies a balance relation. The distinction into different kinds depends on the physical theory. In most everyday situations, the distinction corresponds to different [chemical elements¹](#), and each satisfies its own balance. These balances are the basis of [stoichiometry²](#). If we observe phenomena such as nuclear fission or fusion, however, we notice that the balances of chemical elements are not really satisfied. With such phenomena we make a different distinction of types of matter, for instance *baryons* and *leptons*, and each satisfies again its own balance. It is unclear whether these balances might be broken in other physical phenomena at smaller scales. In these notes we shall usually consider chemical elements as the different kinds of matter, making some exceptions in discussion of nuclear phenomena.

11.1 Electric charge and magnetic flux**11.2 Magnetic flux****11.3 Energy****11.4 Momentum****11.5 Angular momentum****11.6 Entropy**

URLs for chapter 11

1. <https://doi.org/10.1351/goldbook.C01022>
2. <https://doi.org/10.1351/goldbook.S06026>

Chapter 12

Thermodynamics

12.1 Notes on “quasi-static” processes

In general, none of the statements "reversible \Rightarrow quasi-static" or "quasi-static \Rightarrow reversible" is true.

A counterexample to the second implication are systems with internal state variables, which cannot be made non-dissipative, no matter how slowed-down they are. See the discussion and mathematical analysis in Astarita § 2.5.

A counterexample to the first implication is a system of spins in a crystal lattice. It is possible to *reversibly* bring the system from an equilibrium state to another with opposite temperature by reversing the external magnetic field *as fast as possible* – and therefore *not* through a quasi-static process. In fact it is key here that the process be *not-quasi-static*, but as fast as possible, because a slow change of the external magnetic field would lead to an irreversible process with dissipation. For more details see the discussion in Buchdahl, Lecture 20.

The point is that for some systems a *fast* change can actually prevent the onset of dissipative phenomena, and so the process needs to be fast if we want it to be reversible. Adiabatic processes often also need to be fast (as a curious historical fact, Truesdell & Bharatha, Preface p. xii, remark that “In introducing what we today call an ‘adiabatic process’, Laplace called it ‘a sudden compression’, in which he was followed by Carnot”).

In fact, clearly non-quasi-static phenomena such as *explosions* can in some circumstances be described by *reversible* processes! This is possible if the explosion involves many shock waves, as explained by Oppenheim, chap. 1 p. 63:

If there is more than one shock, the losses in available energy are diminished, so that in the limit, with an infinite number of shocks, they become negligible, and the process acquires the character of a thermodynamically optimal, i.e., reversible, change of state. The study of explosion processes reveals that, indeed, they are associated not with one but with a multitude of shocks.

For explosions see also the mathematical analysis by Dunwoody: *Explosion and implosion in a mixture of chemically reacting ideal gases*, where again reversible-process equations are used.

A caveat about reversible and quasi-static associations is given by Erickson (§ 1.2):

Some tend to associate nearly reversible processes with those taking place very slowly – the "quasi-static" processes. This probably stems, at least in part, from experience with classical theories of heat conduction, viscosity, and so on. However, a ball made of silly putty behaves almost reversibly when bounced rapidly and various other high polymers have similar predilections. So, it seems prudent to be open-minded in considering what may be reversible processes for particular systems.

He later discusses (§ 3.1) the case of bars subjected to dead loads, for which we can have reversible processes under sudden jumps in elongation. He concludes (p. 46) that "the sudden jump provides an example of a process that is reversible but not reasonably considered to be quasi-static".

But there's an important question that underlies our discussion: what do we actually mean by "quasi-static"? We need to specify a time scale, otherwise the term is undefined. For example, a geological process (say, tectonic motion) can be considered as quasi-static – or even completely static – on time scales of minutes or days; but it is not quasi-static on time scales of millions of years.

Whether a process is reversible or not, within any tolerance needed, is an experimental question. We can measure any relevant quantities, say pressure p and exchanged heat q , under the process, and compare them with those, p^* and q^* , determined by the equations for a reversible process. We may find for example that at all times

$$\left| \frac{p - p^*}{p^*} \right| < 0.001 , \quad \left| \frac{q - q^*}{q^*} \right| < 0.001$$

and conclude that the process is reversible, if relative discrepancies of 0.1% or less are negligible in our concrete application.

But suppose that someone tells us "if you want the process to be reversible, you must make sure that it is quasi-static". Alright, but how much is "quasi-static"? Is it OK if the piston moves with a speed of 1 cm/s? or is that too much? How about 1 mm/s? – In fact we may find that for some kind of fluid 1 cm/s is absolutely acceptable for the process to be reversible, whereas for another kind of fluid that speed would lead (at the same temperature) to an irreversible process.

You see how this imprecise situation can lead to circular definitions: "if the process is irreversible, then it means it isn't quasi-static" – but then we are actually *defining* "quasi-static" in terms of "reversible"! Any statement of the kind "reversible \Rightarrow quasi-static" or "quasi-static \Rightarrow reversible" then becomes not a matter of experimental verification, but of pure *semantics*. At this point we can simply get rid of "quasi-static" terminology since it doesn't bring any new physics to the table. This circularity is admitted for example by Callen in discussing irreversible gas expansion (Problem 4.2-3 p. 99):

The fact that $dS > 0$ whereas $dQ = 0$ is inconsistent with the presumptive applicability of the relation $dQ = T dS$ to all quasi-static processes. We define (by somewhat circular logic!) the continuous free expansion process as being "essentially irreversible" and *non-quasi-static*.

A similar criticism can be read in Astarita, § 2.9, p. 62, where he also provides a mathematical quantification of quasi-static, similar to the one given above for reversibility:

Often this point is circumvented by bringing in another difficult concept, that of a quasi-static transformation, which proceeds "through a sequence of equilibrium states." Quasi-static is an impressive word, but the only meaning which can be attached to it is the less impressive word "slow" – and how can one speak of slowness without implying the concept of time? How slow is slow enough? If one chooses to develop a thermodynamic theory (rather than a thermostatic one), the answer is easy. For instance, in the case of a system where the state is V, T, \dot{V} [the latter is the rate of change of V], one needs to assume

that [the non-equilibrium pressure] $p(V, T, \dot{V})$ is a Taylor-series expandable at $\dot{V} = 0$ to obtain [that

$$p = p^* + \frac{\partial p}{\partial \dot{V}} \Big|_{\dot{V}=0} \dot{V} + O(\dot{V}^2),$$

where $p^* = p(V, T, 0)$ is the pressure at equilibrium]. One then reaches the conclusion that if the condition

$$\dot{V} \ll \frac{p^*}{\partial p / \partial \dot{V} \Big|_{\dot{V}=0}}$$

is satisfied, then indeed the difference between p and p^* is negligibly small as compared to p^* , and thus the process can be regarded as a quasi-static one.

Criticisms against the fuzzy notion of "quasi-static" have appeared in many other works. Truesdell & Bharatha (Preface p. xii), make the historical remark that "the 'quasi-static process' was barely mentioned for the first time in 1853 and was altogether foreign to the early work [in thermodynamics]". See also the mathematical analysis by Serrin: *On the elementary thermodynamics of quasi-static systems and other remarks*.

I also want to point out that "quasi-static" in some works has specific meanings somewhat unrelated to the discussion above. For example that the rate of increase of the total kinetic energy K of the system is negligible, so that the law of energy balance, which in its full generality is

$$\frac{d(U + K)}{dt} = Q + W$$

(that is, the rate of increase of internal energy U and kinetic energy is equal to the heat rate Q and work rate W provided to the system) can be approximated by

$$\frac{dU}{dt} = Q + W.$$

Or that similar inertial terms in the motion of the system are negligible. See for example the book by Day, chap. 2.

But note that such definitions of "quasi-static" have, again, *no* a-priori relation with reversibility.

Finally, the equation $dS = Q/T$ is only valid for a process that is:

- reversible (by definition),

- closed (no exchange of mass),
- with a homogeneous surface temperature,
- without bulk heating (such as instead happens in a microwave oven).

Under the last three conditions we have in general that $dS \geq Q/T$; when the equality sign is satisfied, then the process is *defined* as reversible. See Astarita, § 1.5, or Müller & Müller, for the different forms of the second law under different circumstances. This equation may be valid in quasi-static and non-quasi-static processes, as explained above.

Bibliography

(“de X” is listed under D, “van X” under V, and so on, regardless of national conventions.)

- Anderson, E. K., Baker, C. J., Bertsche, W., Bhatt, N. M., Bonomi, G., Capra, A., Carli, I., Cesar, C. L., et al. (2023): *Observation of the effect of gravity on the motion of antimatter*. Nature **621**⁷⁹⁸⁰, 716–722. doi:[10.1038/s41586-023-06527-1](https://doi.org/10.1038/s41586-023-06527-1).
- Ashby, N. (2002): *Relativity and the global positioning system*. Phys. Today **55**⁵, 41–47. doi:[10.1063/1.1485583](https://doi.org/10.1063/1.1485583).
- (2003): *Relativity in the global positioning system*. Living Rev. Relativity **6**, 1–42. doi:[10.12942/lrr-2003-1](https://doi.org/10.12942/lrr-2003-1).
- Astarita, G. (1990): *Thermodynamics: An Advanced Textbook for Chemical Engineers*, 2nd pr. (Springer, New York). doi:[10.1007/978-1-4899-0771-4](https://doi.org/10.1007/978-1-4899-0771-4). First publ. 1989.
- Chew, G. F. (1970): *Hadron bootstrap: triumph or frustration?* Phys. Today **23**¹⁰, 23–28. doi:[10.1063/1.3021778](https://doi.org/10.1063/1.3021778).
- Dirac, P. A. M. (1955): *Gauge-invariant formulation of quantum electrodynamics*. Can. J. Phys. **33**¹¹, 650–660. doi:[10.1139/p55-081](https://doi.org/10.1139/p55-081).
- Dyson, F. J. (2007): *Why is Maxwell’s theory so hard to understand?* In: *Antennas and propagation, 2007. eucap 2007*, ed. by The Institution of Engineering and Technology (IEEE, Edinburgh): 1–6. <http://www.damtp.cam.ac.uk/user/tong/em/dyson.pdf>.
- Einstein, A. (1905a): *Zur Elektrodynamik bewegter Körper*. Ann. Phys. (Berl.) **17**, 891–921. Transl. in Einstein (1989a), Doc. 23 pp. 140–171.
- (1905b): *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?* Ann. Phys. (Berl.) **18**, 639–641. Transl. in Einstein (1989a), Doc. 24 pp. 172–174.
- (1989a): *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909. (English translation)*. (Princeton University Press, Princeton). Transl. of Einstein (1989b) by Anna Beck and Peter Havas.
- (1989b): *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909*. (Princeton University Press, Princeton). Ed. by John Stachel. Transl. in Einstein (1989a).
- Faraday, M. (1846): *Thoughts on ray-vibrations*. Philos. Mag. **28**¹⁸⁸, 345–350. doi:[10.1080/1478644608645431](https://doi.org/10.1080/1478644608645431).
- Fliegel, H. F., DiEsposti, R. S. (1996): *GPS and relativity: an engineering overview*. PTTI Proc. **28**, 189–200. https://ui.adsabs.harvard.edu/link_gateway/1997ptti.conf..189F/ADS_PDF, <https://apps.dtic.mil/sti/citations/ADA516975>.

- Galilei, G. (1623): *Il Saggiatore*. http://www.liberliber.it/biblioteca/g/galilei/il_saggiatore/html/index.htm; parts transl. in Seeger (1966).
- iso (2009): ISO 80000-1:2009: *Quantities and units 1: General*. International Organization for Standardization.
- (2019): ISO 80000-2:2019: *Quantities and units 2: Mathematics*. International Organization for Standardization.
- Kandus, A., Tsagas, C. G. (2008): *Generalized Ohm’s law for relativistic plasmas*. Mon. Not. R. Astron. Soc. **385**², 883–892. doi:[10.1111/j.1365-2966.2008.12862.x](https://doi.org/10.1111/j.1365-2966.2008.12862.x).
- Kovalevsky, J., Seidelmann, P. K. (2004): *Fundamentals of Astrometry*. (Cambridge University Press, Cambridge). doi:[10.1017/CBO9781139106832](https://doi.org/10.1017/CBO9781139106832).
- Maxwell (Clerk Maxwell), J. (1855): *On Faraday’s lines of force*. Trans. Cambridge Philos. Soc. **10**, 27–83. First read 1855–1856; repr. in Maxwell (2010a) doc. VIII pp. 155–229.
- (1870): *Mathematics and physics*. Rep. Br. Assoc. Adv. Sci. **40**, 1–9. <https://archive.org/details/reportofbritishab71brit>. Repr. in Maxwell (2010b) doc. XLI pp. 215–229. See also Dyson (2007).
 - (1878): Tait’s “*Thermodynamics*”. Nature **17**^{431, 432}, 257–259, 278–280. doi:[10.1038/017257a0](https://doi.org/10.1038/017257a0). Repr. in Maxwell (2010b) doc. XCI pp. 660–671.
 - (2010a): *The Scientific Papers of James Clerk Maxwell*. Vol. 1, repr. (Cambridge University Press, Cambridge). Ed. by W. D. Niven. doi:[10.1017/CBO9780511698095](https://doi.org/10.1017/CBO9780511698095), <https://archive.org/details/scientificpapers01maxwuoft>. First publ. 1890.
 - (2010b): *The Scientific Papers of James Clerk Maxwell*. Vol. 2, repr. (Cambridge University Press, New York). Ed. by W. D. Niven. doi:[10.1017/CBO9780511710377](https://doi.org/10.1017/CBO9780511710377), <https://archive.org/details/scientificpapers02maxwuoft>. First publ. 1890.
- Misner, C. W., Thorne, K. S., Wheeler, J. A. (2017): *Gravitation*, repr. (Princeton University Press, Princeton and Oxford). With a new foreword by David I. Kaiser and a new preface by Charles W. Misner and Kip S. Thorne. First publ. 1970. <https://archive.org/details/GravitationMisnerThorneWheeler>.
- Müller, J., Soffel, M., Klioner, S. A. (2008): *Geodesy and relativity*. J. Geod. **82**³, 133–145. doi:[10.1007/s00190-007-0168-7](https://doi.org/10.1007/s00190-007-0168-7).
- Newton, I. (1726): *Philosophiae naturalis principia mathematica*, “Tertia aucta & emendata” ed. (Guil. & Joh. Innys, London). <http://archive.org/details/principiareprint00newtuoft>, <http://archive.org/details/principia00newtuoft>. First publ. 1687; transl. in Newton (1846; 1974a,b).
- (1846): *Newton’s Principia: The Mathematical Principles of Natural Philosophy*, first American, “carefully rev. and corr.” ed. (Daniel Ade, New York). Transl. of Newton (1726) by Andrew Motte. With Newton’s system of the world and a life of the author by N. W. Chittenden.
 - (1974a): *Sir Isaac Newton’s Mathematical Principles of Natural Philosophy and his system of the world. Vol. One: The Motion of Bodies*. (University of California Press, Berkeley). Transl. of Newton (1726) by Andrew Motte, rev. and supplied with an historical and explanatory appendix by Florian Cajori.
 - (1974b): *Sir Isaac Newton’s Mathematical Principles of Natural Philosophy and his system of the world. Vol. Two: The System of the World*. (University of California Press, Berkeley). Transl. of Newton (1726) by Andrew Motte, rev. and supplied with an historical and explanatory appendix by Florian Cajori.

- Parker, E. N. (1974a): *Hydraulic concentration of magnetic fields in the solar photosphere. II. Bernoulli effect*. *Astrophys. J.* **190**, 429–436. doi:[10.1086/152894](https://doi.org/10.1086/152894). See also Parker (1974b).
- (1974b): *Hydraulic concentration of magnetic fields in the solar photosphere. I. Turbulent pumping*. *Astrophys. J.* **189**, 563–568. doi:[10.1086/152835](https://doi.org/10.1086/152835). See also Parker (1974a).
- Petit, G., Wolf, P. (2005): *Relativistic theory for time comparisons: a review*. *Metrologia* **42**³, S138–S144. doi:[10.1088/0026-1394/42/3/S14](https://doi.org/10.1088/0026-1394/42/3/S14), http://geodesy.unr.edu/hanspeter_plag/library/geodesy/time/met5_3_S14.pdf.
- Ryutova, M. (2018): *Physics of Magnetic Flux Tubes*, 2nd ed. (Springer, Heidelberg). doi: [10.1007/978-3-319-96361-7](https://doi.org/10.1007/978-3-319-96361-7). First publ. 2015.
- Seeger, R. J., ed. (1966): *Galileo Galilei, His life and His Works*. (Pergamon, Oxford).
- Styer, D. F. (2000): *Insight into entropy*. *Am. J. Phys.* **68**¹², 1090–1096. doi:[10.1119/1.1287353](https://doi.org/10.1119/1.1287353), <http://www.nd.edu/~powers/ame.20231/styer2000.pdf>.
- Taylor, E. F., Wheeler, J. A. (2000): *Exploring Black Holes: Introduction to General Relativity*. (Addison Wesley Longman, San Francisco).
- Truesdell III, C. A. (1964): *Die Entwicklung des Drallsatzes*. *Z. Angew. Math. Mech.* **44**^{4/5}, 149–158. First delivered as a lecture 1963. See also Truesdell (1968a).
- (1966): *Six Lectures on Modern Natural Philosophy*. (Springer, Berlin). doi:[10.1007/978-3-662-29756-8](https://doi.org/10.1007/978-3-662-29756-8).
- (1968a): *Whence the law of moment of momentum?* In: Truesdell (1968b): ch. V:pp. 239–271. doi:[10.1007/978-3-642-86647-0_5](https://doi.org/10.1007/978-3-642-86647-0_5). First delivered as a lecture 1963. See also Truesdell (1964).
- (1968b): *Essays in the History of Mechanics*. (Springer, Berlin). doi:[10.1007/978-3-642-86647-0](https://doi.org/10.1007/978-3-642-86647-0).
- (1987): *An Idiot’s Fugitive Essays on Science: Methods, Criticism, Training, Circumstances*, 2nd pr., rev. and augmented. (Springer, New York). First publ. 1984.
- Walpole, S. C., Prieto-Merino, D., Edwards, P., Cleland, J., Stevens, G., Roberts, I. (2012): *The weight of nations: an estimation of adult human biomass*. *BMC Public Health* **12**, 439. doi:[10.1186/1471-2458-12-439](https://doi.org/10.1186/1471-2458-12-439).
- Whittaker, E. T. (1951): *A History of the Theories of Aether and Electricity. Vol. 1: The Classical Theories*, rev. and enl. ed. (Thomas Nelson and Sons, London). <https://archive.org/details/e-t-whittaker-a-history-of-the-theories-of-aether-and-electricity-vol-1-london-n>. First publ. 1910.
- Wojtan, C., Thürey, N., Gross, M., Turk, G. (2009): *Deforming meshes that split and merge*. *ACM Trans. Graph.* **28**³, 1–10. doi:[10.1145/1531326.1531382](https://doi.org/10.1145/1531326.1531382).