

# 2

## Essential Quantum Features

The underlying message to take away from this chapter is that nature works very differently at small scales than we would expect from our everyday experiences.

We will start with a short discussion of the three most important features of quantum mechanics:

1. **We need waves to describe particles.**
2. **Some physical properties are quantized.**
3. **Quantum systems have an inherent uncertainty.**

Afterward, we will talk about how we can describe these unintuitive features using an intuitive physical framework.

### 2.1 The Heart of Quantum Mechanics

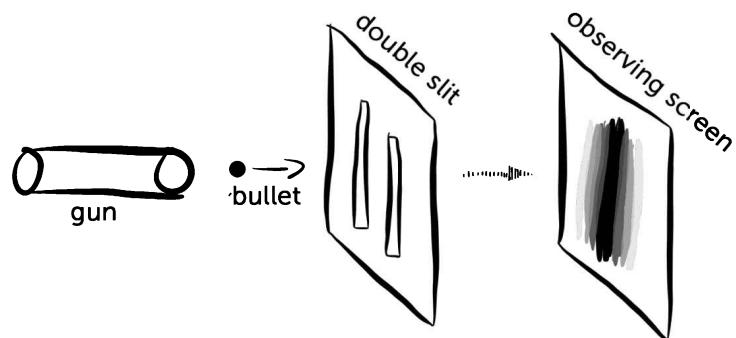
Richard Feynman once remarked in his lectures that one experiment "has in it the heart of quantum mechanics. In reality, it contains the only mystery".<sup>1</sup>

<sup>1</sup> Richard Feynman. *The Feynman lectures on physics*. Basic Books, a member of the Perseus Books Group, New York, 2011. ISBN 978-0465025015

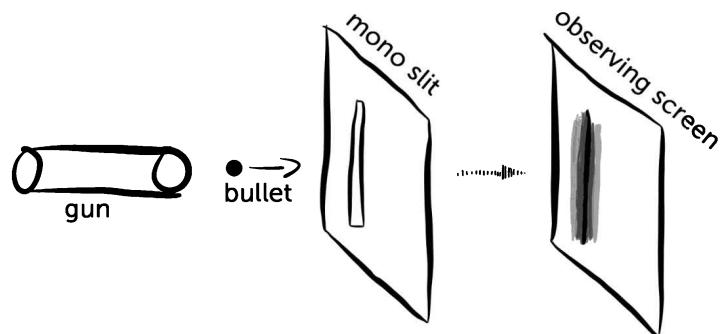
The experiment he was referencing is astonishingly straightforward. All we need is a beam of electrons, a wall with two little slits in it, and a screen behind the wall that detects the electrons. This experimental setup is known as the **double slit experiment**.

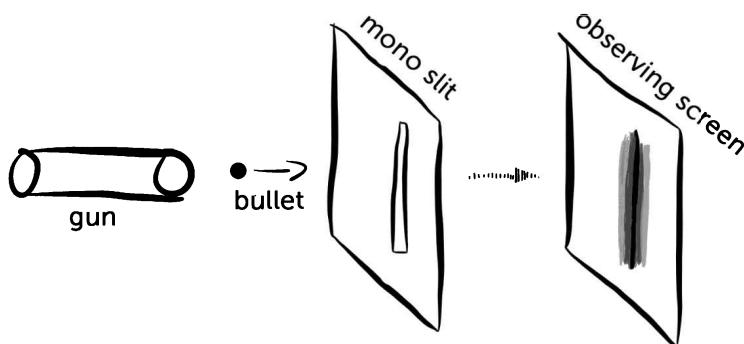
Before we discuss the experiment with actual electrons, there are a few preliminary things that we should talk about.

First, if we shoot a beam of bullets (rather than electrons) towards a double slit and observe the pattern behind it, the results is rather dull:



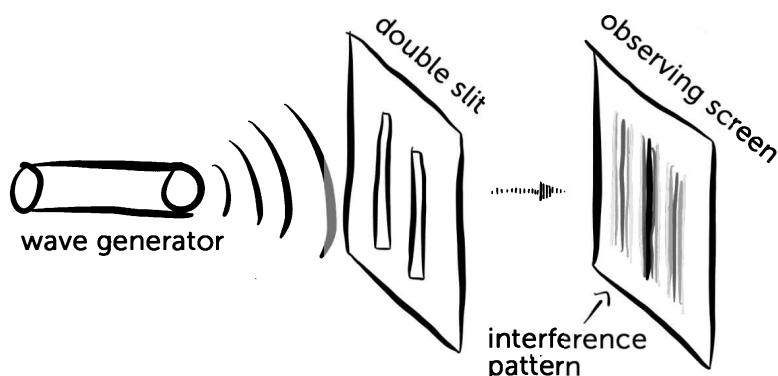
Most bullets end up in the middle of the screen and progressively fewer if we move away from the middle. We can understand this pattern by considering the same experiment with just one slit:





The result of the double-slit experiment is simply a sum of the results obtained from each single-slit experiment.

If we replace the bullet with a wave (e.g., a water or sound wave), the result shown on the observing screen is much more interesting:



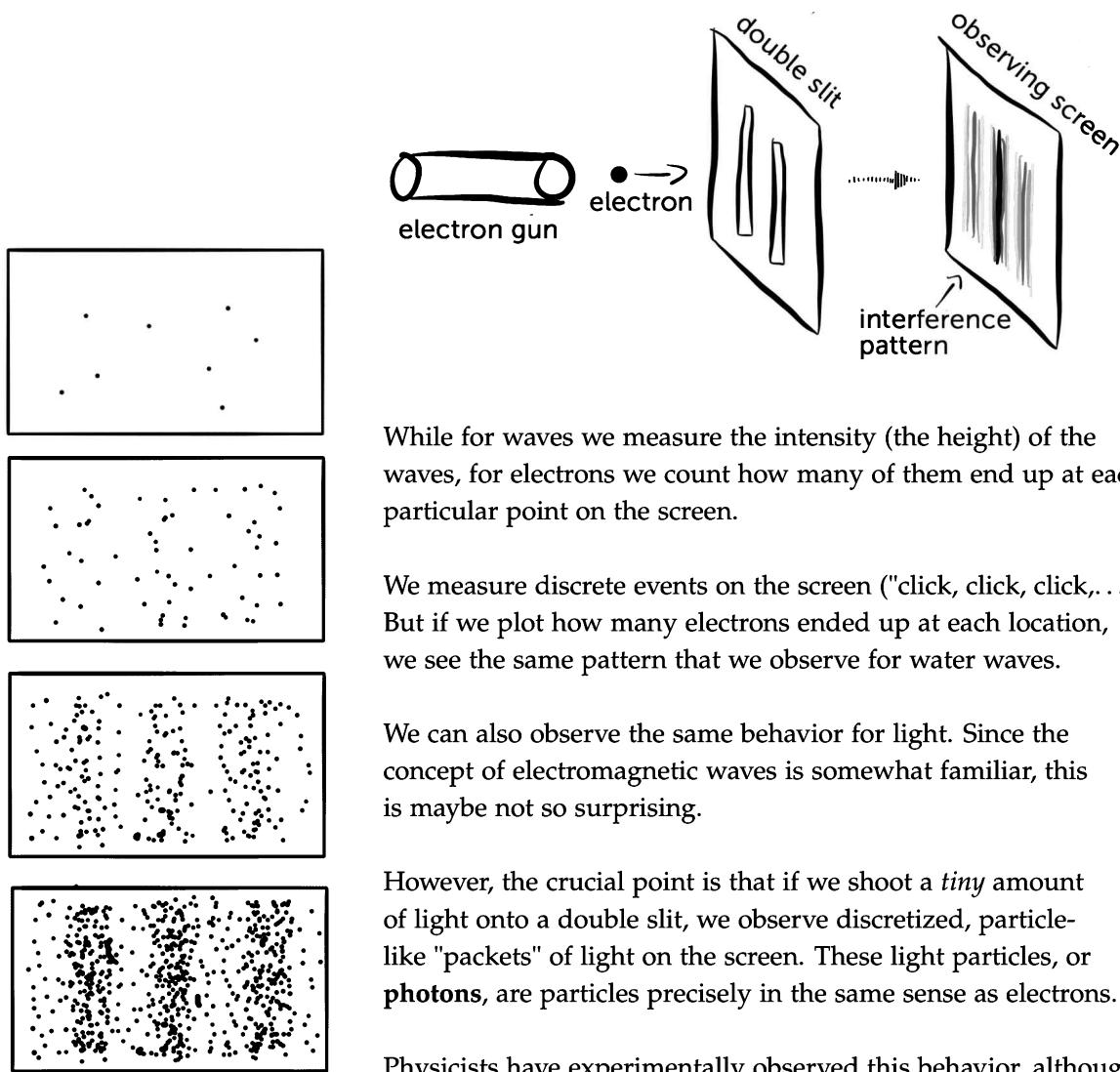
The pattern we see on the screen is known as an **interference pattern**. This pattern arises because the wave goes through both slits and we can imagine that at each slit a new wave originates. Behind the double slit, the new waves propagate and overlap, resulting in the interference pattern observed on the screen.

So, what does all this have to do with quantum mechanics?

Well, we usually think of particles like electrons as analogous to very tiny bullets. Therefore, when we shoot electrons towards

a double slit, we would expect results similar to those obtained using bullets.

The crazy thing is that if we perform the double slit experiment with electrons, we actually see an interference pattern on the screen!



While for waves we measure the intensity (the height) of the waves, for electrons we count how many of them end up at each particular point on the screen.

We measure discrete events on the screen ("click, click, click,..."). But if we plot how many electrons ended up at each location, we see the same pattern that we observe for water waves.

We can also observe the same behavior for light. Since the concept of electromagnetic waves is somewhat familiar, this is maybe not so surprising.

However, the crucial point is that if we shoot a *tiny* amount of light onto a double slit, we observe discretized, particle-like "packets" of light on the screen. These light particles, or **photons**, are particles precisely in the same sense as electrons.

Physicists have experimentally observed this behavior, although we still struggle to *understand* it. There is no problem, however, in *describing* the phenomenon and this is what quantum mechanics is primarily about.<sup>2</sup>

<sup>2</sup> We will talk a bit about how we possibly can interpret this crazy behavior in Chapter 15.

So the bottom line is:<sup>3</sup>

**While electrons as well as other elementary particle are *discrete* objects, we need waves to describe them.**

Before we move on, I want to warn you about one thing. Many textbooks include much discussion of things like the mysterious wave-particle duality. Many people unfamiliar with quantum mechanics may wonder how an electron could be a particle and a wave at the same time.

Please ignore this kind of idle speculation. The situation is not as crazy as some would lead you to believe. Electrons, photons, and all other elementary particles are *particles*. Period. This is what every experiment tells us. Our detectors make "click, click, click".

Waves are merely one convenient *mathematical tool* for describing the behavior of these particles.<sup>4</sup> Since most people have experience with waves, using them as a mathematical tool allows us to develop some intuition for the behavior of quantum systems.

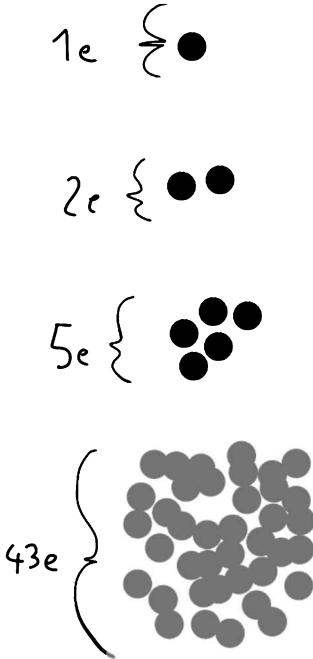
However, this wave description is only a mathematical tool. It is also possible to describe everything in quantum mechanics completely without using waves at all.<sup>5</sup> So please, don't let yourself get confused by such discussions.

In the following two sections we will talk about the two most important features of quantum mechanics. We can directly understand both when we use waves to describe particles.

<sup>3</sup> This is not 100% correct. Waves are the simplest way to get a first grasp on what quantum mechanics is all about. We will discuss in Section 12.2 how we can describe quantum behaviour without waves.

<sup>4</sup> These waves were historically called "matter waves", a term introduced by French physicist Louis de Broglie.

<sup>5</sup> We will talk about this in detail in Chapter 14.



<sup>7</sup> A special type of speculative models known as "grand unified theories" has been suggested as a solution to this conundrum.

## 2.2 What Does "Quantum" Mean?

In classical mechanics, we can measure for basic quantities like energy or angular momentum any value we can imagine. This is not always the case for quantum mechanical systems. In many quantum systems, basic quantities only arise in discrete chunks, or **quanta**. As such, we say that these properties are **quantized**.

A famous example is electric charge. Any measurement of electric charge yields an integer multiple of the basic charge  $e$ :

$$Q = N \times e,$$

where  $N$  is an integer and  $e$  is the absolute value of the charge of an electron.<sup>6</sup> In particular, this means that a measurement of electric charge of some object in nature possibly yields  $2e$  or  $977e$ , but never something like  $2.12e$  or  $34.76e$ . In this sense, electric charge is quantized. Why this happens is a different question and, unfortunately, we don't know the answer yet.<sup>7</sup> But this is what makes quantum mechanics so interesting. There are still lots of fundamental things we do not understand.

No theoretical physicist predicted the quantization of electric charge. It was also not immediately discovered when physicists were developing electrodynamics because the charges of macroscopic objects (e.g., a metal sphere) are typically on the order of  $Q_{\text{sphere}} \approx 4.806 \text{ C}$ . Compared to the charge of a single electron, these macroscopic objects have a huge amount of charge.

As a result, it seems as if we can produce spheres with arbitrary charges. This is an illusion. If we measure the electric charge of a macroscopic object very, very precisely, we notice that even such macroscopic charges are integer multiples of  $e$ .<sup>8</sup>

$$\begin{aligned} Q_{\text{sphere1}} &\approx 4.806 \text{ C} \approx 3000 \cdot 10^{16}e \\ Q_{\text{sphere2}} &\approx 5.01 \text{ C} \approx 3127 \cdot 10^{16}e. \end{aligned} \quad (2.1)$$

The fundamentally discrete chunks of electric charge are tiny compared to the charges of macroscopic objects. When we change the charge of a sphere, we get a discrete sequence like

$$Q_{sphere1} = 4.80652986239999955569404683720 \text{ C}$$

$$Q_{sphere1+1e} = 4.80652986239999955585426449928 \text{ C}$$

$$Q_{sphere1+2e} = 4.806529862399999556014482161367 \text{ C}$$

$$Q_{sphere1+3e} = 4.80652986239999955633491748552 \text{ C}$$

:

$$Q_{sphere1+2948745599054545e} = 4.80700230352594876390867348334 \text{ C}$$

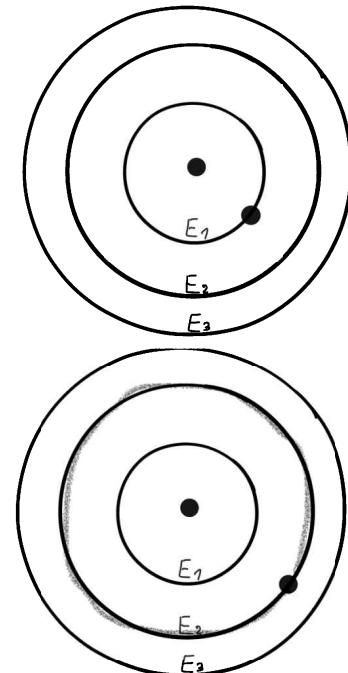
If we are only concerned with the first four digits, we can conclude that we can produce any value of the charge that we like. It's only when we measure the charge *extremely* precisely that we notice that changes occur in discrete steps.<sup>9</sup>

The same thing also happens for other fundamental physical quantities. In many quantum systems, we can no longer measure any arbitrary value for the energy or angular momentum.<sup>10</sup> Instead, they *only* appear in discrete chunks and are therefore quantized.

A famous example of quantization can be found in the discrete energy levels populated by the electrons in an atom. Nothing in classical electrodynamics tells us otherwise. In quantum mechanics, however, only a discrete set of energy levels is allowed. We can observe this by measuring the radiation emitted whenever an electron jumps from one energy level to a lower one. The energy of the radiation is precisely the energy difference between the energy levels. The energy of the electron does not change continuously from one value to another. Instead, it jumps directly. This is where the notion of **quantum jumps** comes from.

<sup>9</sup> We can understand this quantization since when we charge a sphere we add or remove electrons from it. Each additional electron contributes an additional charge  $e$ . This is why we have discrete steps. However, a mystery is why, for example, protons do not have some random value of the charge, but exactly the charge  $-e$ .

<sup>10</sup> In speculative theories like loop quantum gravity, even space and time are quantized.

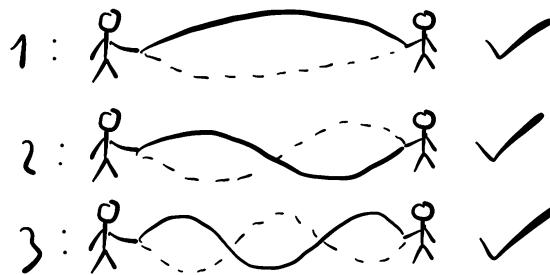



---

Okay, what does all this have to do with the fact that we need waves to describe particles?

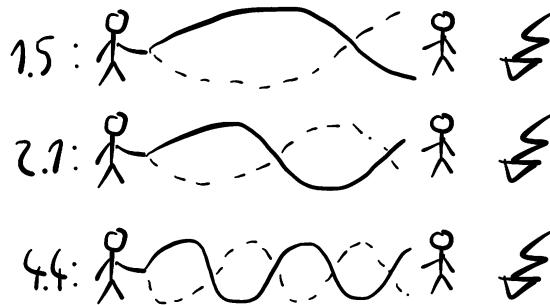
Well, waves naturally quantize physical properties!

To understand this, let's consider a rope which we hold under constant tension:



No matter how the two hands try to make the rope vibrate, the rope will only vibrate with a quantized set of modes. The two hands fix the rope at both ends. As a result of this constraint, the rope can only vibrate with fixed modes.

It is physically impossible to make the rope vibrate with a mode outside of this fixed set.

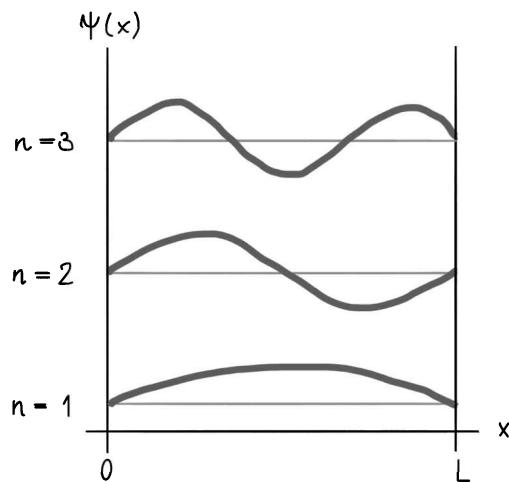


The same thing also happens when we describe particles using waves. For example, let's consider an electron confined in a box.<sup>11</sup>

If we describe the electron using a wave, we notice that only very particular wave shapes are allowed:<sup>12</sup>

<sup>11</sup> We will discuss this example in detail in Chapter 7.

<sup>12</sup> This is a result of the boundary conditions imposed by the box and is completely analogous to what we have discussed for the rope example.



The **frequency** of the wave associated with a given particle is directly related to its energy as<sup>13</sup>

$$\nu = \frac{E}{h}, \quad (2.2)$$

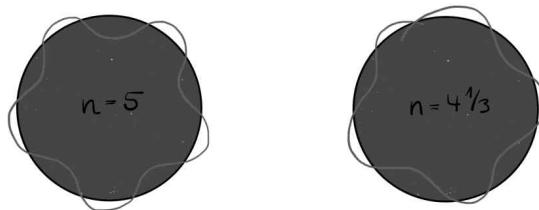
where  $h$  is a constant known as the Planck constant,  $E$  is the energy of the particle and  $\nu$  is the frequency of the corresponding wave<sup>14</sup>. This means that our particle in a box can't have arbitrary energy values. Since only a discrete set of modes is allowed, the particle can only take on a discrete set of energies. This is how energy becomes quantized in quantum mechanics!

Similarly, we can understand why there are only discrete energy levels in atoms. As a naive picture, imagine electrons orbiting the nucleus. Classically, any orbit is allowed. However, as soon as we describe electrons with waves, this is no longer the case. The electron "orbits" correspond to **standing waves** around the nucleus. For a fixed radius  $r$ , we are effectively dealing with something like a box<sup>15</sup>. The wave must coincide at the endpoint and the starting point, therefore allowing only specific wave modes:

<sup>13</sup> We will discuss the origin of this formula later. For the moment, we are only interested in the basic message and want to build some intuition.

<sup>14</sup> In slogan form: The higher the frequency, the higher the energy.

<sup>15</sup> Imagine in the picture of the box above that the points  $x = 0$  and  $x = L$  coincide. The fixed radius  $r$  is known as the **Bohr radius**. Physically, the Bohr radius represents the most statistically likely distance between the nucleus and the electron.



<sup>16</sup>This explanation is known as the **Bohr model**. Be warned that this slightly outdated model is not entirely correct, but works well for the hydrogen atom. The details in a real atom are, of course, a lot messier. However, the Bohr model still gives a nice intuitive feeling of how quantization happens in quantum mechanics.

<sup>17</sup>Take note that the energy of a *free* particle, for example, is not quantized. The energy is only quantized for bounded systems like a box. However, the electric charge is always quantized. We will discuss this in a bit more detail in Chapter 5.

This is why atomic energy levels are quantized.<sup>16</sup>

To summarize: Many familiar quantities like energy are not always continuous. In certain quantum systems, they only appear in discrete chunks.<sup>17</sup>

Whenever this is true, we say the corresponding quantity is quantized. This fact is not observable in everyday life because the intervals between allowed values are tiny. Quantization is a consequence of the "wave nature" of particles.

## 2.3 Uncertainty

A second crucial feature of quantum mechanics is that there is an *intrinsic* uncertainty.

A somewhat naive (but helpful) explanation of quantum uncertainty goes as follows:<sup>18</sup>

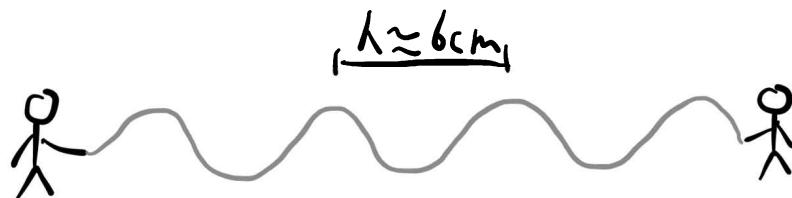
Imagine that we want to measure the position of an electron by taking a picture of it. To take this picture, we need to bounce light off the electron's "surface", since without light we can't see the electron. However, the light we bounced off of the electron unavoidably caused the electron to move a bit. As a result, the momentum of the electron is changed each time we measure its position. What this means is that we can't know the position and momentum of an electron at the same time with arbitrary precision. We have to decide which property we want to know exactly. We can either measure the position exactly or measure the momentum exactly, but we cannot measure both at the same time. There is a limit to how much we can know about an electron.

<sup>18</sup> The quantum uncertainty is often also called **Heisenberg uncertainty** and the following explanation is actually how Werner Heisenberg liked to explain it.

Again you may wonder: What does this have to do with our wave description of particles?

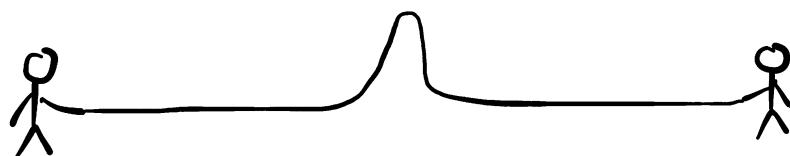
In the previous section, we saw why quantization is a direct result of this new description. Now we will see that quantum uncertainty is also a *direct* consequence of the "wave nature" of particles!

Once more, let's consider a rope. We can generate a wave in a long rope by shaking it rhythmically up and down:



If someone were to ask us where the wave is, we wouldn't have a good answer since the wave is spread out. But if we get asked: "What's the wavelength of the wave?", we could easily answer this question and state: "It's around 6cm."

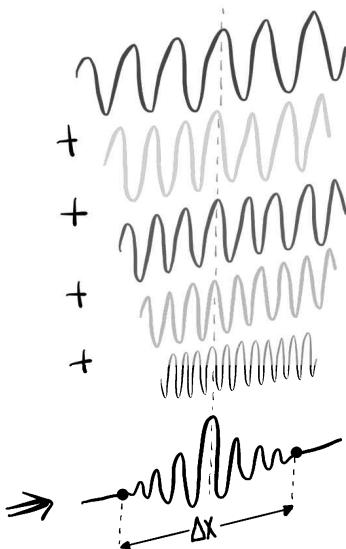
We can also generate a different kind of wave in a rope by jerking it only once.



This way, we get a narrow bump that travels down the line. For this kind of wave, we can easily answer the question: "Where precisely is the wave?". But we have a hard time answering the question: "What's the wavelength of this wave?", since the wave isn't periodic and it's completely unclear how (or if) we can assign a wavelength to it.

Analogously, we can generate any kind of wave in between these two edge cases and there is always a trade-off. The more precisely the position of the wave is localized, the more ambiguous the wavelength becomes, and vice versa.

To make this idea more precise, we can think of a localized wave as a **superposition** of dozens of other waves with well-defined wave-lengths<sup>19</sup>.



<sup>19</sup> Such waves with well-defined wavelengths are known as **plane waves**. The expansion of a general bump in terms of such plane waves is exactly the idea behind the Fourier transform. For more information on this see Appendix B. The uncertainty we end up with this way is a general feature of waves and known as the **bandwidth theorem**.

If we add lots of waves with different wavelengths, they will average out almost everywhere. But we can arrange the waves such that, in a small region, they don't cancel each other out.

This is true for all waves. Since in quantum mechanics, we describe particles using waves, it also applies here.

In quantum mechanics, the **wavelength** is directly related to its

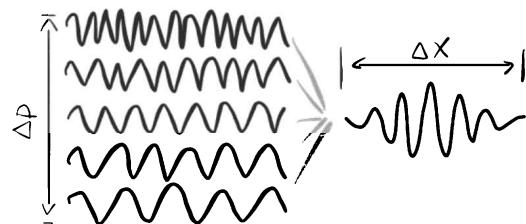
momentum

$$\lambda = \frac{h}{p} \quad (2.3)$$

The larger the momentum  $p$ , the smaller the wavelength  $\lambda$  of the wave that describes the particle. Therefore, a spread in wavelength corresponds to a spread in momentum.<sup>20</sup>

What this means in physical terms is exactly what we talked about above: We can't know the location and momentum of particles with arbitrary precision:

<sup>20</sup> Once more we use a formula without deriving it. We do this to develop some intuitive understanding of quantum waves first and postpone the actual derivation to a later chapter.



The uncertainty relation that we arrive at tells us:

**The more precisely we determine the location of a particle, the less precisely we are able to determine its momentum and vice versa.**

---

Before we move on, let's summarize what we have learned so far.

The double-slit experiment tells us that we need *waves* to describe *particles*. This "wave nature" of particles then leads us to two essential features of quantum systems: quantization and uncertainty.

Since only special waveforms are allowed for a particle in a box, we end up with a discrete set of allowed states. In physical terms, this means that only a discrete set of energy values is allowed for a particle in a box. As a result, we say that the energy of this system is quantized.

Moreover, we have learned that for waves there is always a trade-off between the localization of the wave and how well we can determine its wavelength.

We can think of a localized wave bump as a sum of many waves with well-defined wavelengths. In other words, a localized waveform corresponds to a superposition of waves with different wavelengths and, therefore, not to one specific wavelength. In turn, a waveform with well-defined wavelength (a plane wave) is not localized anywhere but spreads out all over space.

In quantum mechanics, there is a direct relationship between wavelength and momentum. In addition, the localization of the wave encodes information about the location of the particle in question.

Taken together, this tells us that (thanks to the "wave nature" of particles) we can't simultaneously measure both the location and momentum of a particle with arbitrary precision.

---

Now it's time to develop a framework that will allow us to describe all these weird features of quantum systems.

# 3

## The Quantum Framework

We will start this chapter by developing a general framework to describe nature. Afterward, we will see how easily we can describe quantum mechanics using this general framework.

The most important ingredient that we need is *something* that describes our physical system in question. We invent a new mathematical symbol that does the job:  $|\Psi\rangle$ .

Everything about the physical system is encoded in  $|\Psi\rangle$ . The standard name for this type of object is a **state vector** or **ket**<sup>1</sup>.

Each different possible state of our system corresponds to a different ket  $|\Psi_1\rangle, |\Psi_2\rangle, \dots$

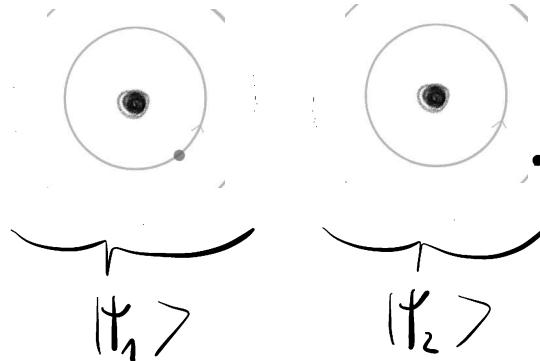
Think about an atom. An electron surrounding a nucleus can have different energies. It can be in the ground state (the state with the lowest energy), but it can also be in an excited state.

In our framework, for example,

- ▷  $|\Psi_1\rangle$  denotes the configuration with the electron in the ground state and

<sup>1</sup>Okay admittedly, this sounds horribly abstract. But I promise this notation is clever and will make a lot of sense in a moment. In particular, the name "ket" will make more sense in a moment.

▷  $|\Psi_2\rangle$  corresponds to the configuration with the electron in the next highest excited state.



These are simply labels for different states. For most systems, there are infinitely many possible states. For example, an electron at position  $x = 0$  is a different state than the same electron at position  $x = 0.1$ . Since there are infinitely many possible locations, there are infinitely many states for a free electron. Again, this may sound a bit scary, but we will see in a moment how to deal with this situation.

So far, so good. But how are these abstract objects related to anything we can measure in experiments?

There are only a few ingredients missing. The next thing that we need is a way to extract information from our kets. Suppose we want to know the momentum of our system. To get this information from our ket  $|\Psi_1\rangle$ , we "ask" it:

$$\hat{p} |\Psi_1\rangle = ? \quad (3.1)$$

The object  $\hat{p}$  is called the **momentum operator**. Again, we simply invented a mathematical symbol that does the job. The *operator*  $\hat{p}$  acts on our ket and yields the momentum of our system:

$$\hat{p} |\Psi_1\rangle = p_1 |\Psi_1\rangle . \quad (3.2)$$

Here,  $p_1 \in \mathbb{R}$  is the momentum of the system in the configuration described by  $|\Psi_1\rangle$ . If we ask the same question when the

system is in a different configuration, then we get a different answer:

$$\hat{p} |\Psi_2\rangle = p_2 |\Psi_2\rangle . \quad (3.3)$$

Yes, it's that simple! We act with  $\hat{p}$  on the object that describes our system,  $|\Psi\rangle$ , and get back the momentum that we can measure in an experiment.

Admittedly, the situation is not always *that* simple. In the previous section, we talked about uncertainty — one of the most important features of quantum mechanics. The presence of this uncertainty means that we won't always get answers that are as simple as obtained in the above examples. Instead, there is always some variance. Even if we manage to prepare our system perfectly such that it *definitely* has momentum  $p_1$ , measurements of the system's position yield a range of values. Therefore, we need to add something to the framework that allows us to deal with uncertainty.

But first, what exactly does uncertainty mean?

Well, in the example we discussed above, it means that measurements of the momentum will not always yield  $p_1$ . Given a set of equally prepared atoms, we will sometimes measure the value  $p_1$  and other times the value  $p_2$ . There is a variance in our measurements.

How can we implement this in our framework? Naively, we write

$$|\Psi\rangle = a |\Psi_1\rangle + b |\Psi_2\rangle , \quad (3.4)$$

where  $a$  and  $b$  are numbers that encode how often we measure  $p_1$  and  $p_2$ , respectively. If we act with  $\hat{p}$  on this new ket, we get an answer we don't immediately understand:

$$\begin{aligned} \hat{p} |\Psi\rangle &\stackrel{\text{Eq. 3.4}}{=} \hat{p} (a |\Psi_1\rangle + b |\Psi_2\rangle) \\ &= a \hat{p} |\Psi_1\rangle + b \hat{p} |\Psi_2\rangle \\ &\quad \swarrow \quad \curvearrowright \quad \hat{p} |\Psi_i\rangle = p_i |\Psi_i\rangle, \text{ Eq. 3.2} \\ &= ap_1 |\Psi_1\rangle + bp_2 |\Psi_2\rangle . \end{aligned} \quad (3.5)$$

We can only understand a result like this in a *statistical* sense. This is what the quantum uncertainty demands of us. Quantum mechanics results often only make sense from a statistical perspective. Therefore, we need to add statistical tools to our framework.

Before we can do this, we need to talk for a moment about statistical tools in general<sup>2</sup>.

<sup>2</sup> If you are already familiar with **expectation values** and **standard deviations**, feel free to skip the following section.

### 3.1 Intermezzo: Essential Statistics

One of the simplest but at the same time most important statistical tools is the so-called **expectation value**<sup>3</sup>. The expectation value of a quantity is the average value that we expect when we repeat a given experiment many times. We use it whenever we are forced to make probabilistic judgments<sup>4</sup>.

To understand what an expectation value is, imagine the following situation:

A friend offers to play a game. She flips a coin. If it lands on tails, she pays you \$1.50. But if it lands on heads, you have to pay her \$1. By calculating the expectation value for this system, you can decide whether you should agree to play this game.

- ▷ The probability that a coin lands on heads is  $p_1 = 50\%$ . In this case, the outcome for you is:  $x_1 = -\$1$ .
- ▷ Equally, the probability that a coin lands on tails is  $p_2 = 50\%$ . In this case, the outcome for you is:  $x_2 = +\$1.50$ .

The expectation value is defined as the sum over each outcome times the probability of the outcome

$$\begin{aligned}
 \text{expectation value} &\equiv \sum_{i=1}^2 x_i P_i \\
 &= x_1 P_1 + x_2 P_2 \\
 &= -\$1 \times 50\% + \$1.50 \times 50\% \\
 &= \$0.25. \tag{3.6}
 \end{aligned}$$

So if we play this game many times, we will make a profit. On average we make \$0.25 per game. For example, let's say you play the game only twice, win once, and lose the second time. For the win in the first game, you get \$1.50. For the loss in the second game, you lose \$1. In total, you have therefore won \$0.50 in two games. This equals \$0.25 per game. Of course, here the

<sup>3</sup> Alternative names for the expectation value are: expected value, mean value, or simply the average. However, in the context of quantum mechanics, it is conventional to use the word "expectation value".

<sup>4</sup> Think flipping a coin or tossing a die.

system is so simple that we could have guessed this without calculating anything. But for more complicated systems, the situation quickly gets messy. Quantum systems, in particular, have many possible outcomes with potentially different probabilities of occurring. This makes it much more difficult to predict anything without referring to something like the expectation value.

*To summarize:* the expectation value is the expected average outcome of repeated probabilistic events. A common mathematical notation for the expectation value of a quantity  $x$  is  $\langle x \rangle$ .<sup>5</sup>

There is one additional statistical notion that we need all the time in quantum mechanics. It tells us how much the system, on average, *deviates* from the expectation value. In other words, this notion tells us how much our measurements spread out. If it is zero, we measure the same value all the time. If it is non-zero, it's possible to measure different values. The notion I'm talking about is called the **standard deviation**. It tells us exactly how much our measurements (on average) spread around the expectation value:<sup>6</sup>

$$\Delta x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}. \quad (3.7)$$

The second term under the square root is simply the expectation value squared. For the first term, we again calculate an expectation value, but this time we square each possible outcome before we weigh it with the corresponding probability

$$\langle x^2 \rangle = \sum_i x_i^2 P_i. \quad (3.8)$$

Compare this with

$$\langle x \rangle^2 = \left( \sum_i x_i P_i \right)^2. \quad (3.9)$$

Yes, it makes a difference whether we square the whole sum or each term of the sum!<sup>17</sup>

To understand this definition, let's return to the example from

<sup>5</sup> Maybe you already noticed the similarity of this notation to our previous notation for the kets  $|\Psi\rangle$ . If not, don't worry, because the similarity will become a lot more obvious in a moment.

<sup>6</sup> This strange looking definition will make sense in a moment. And as an aside: The same definition without the square-root is known as **variance**.

<sup>7</sup> For example, for  $x = \{3, -5, 9\}$  we have

$$\begin{aligned} \sum_i x_i^2 &= (3)^2 + (-5)^2 + (9)^2 \\ &= 115 \end{aligned}$$

whereas

$$\begin{aligned} \left( \sum_i x_i \right)^2 &= (3 - 5 + 9)^2 \\ &= 49. \end{aligned}$$

above. The standard deviation for our coin flipping game reads

$$\begin{aligned}
 \Delta x &\stackrel{\text{Eq. 3.7}}{=} \sqrt{\langle x^2 \rangle - \langle x \rangle^2} \\
 &= \sqrt{\sum_i x_i^2 P_i - \left( \sum_i x_i P_i \right)^2} && \curvearrowright \text{Eq. 3.8 and Eq. 3.9} \\
 &= \sqrt{\sum_i x_i^2 P_i - (\$0.25)^2} && \curvearrowright \text{Eq. 3.6} \\
 &= \sqrt{((-\$1)^2 \times 0.5 + (\$1.5)^2 \times 0.5) - (\$0.25)^2} \\
 &= \$\sqrt{1.625 - 0.25^2} \approx \$1.25.
 \end{aligned}$$

There are two important things to take note of:

- ▷ The result isn't zero. As I already emphasized above, it makes a difference whether we square each term in the sum or the complete sum. Otherwise, the standard deviation would always be zero.
- ▷ If we only know that the expectation value of some game is \$0.25, then we are missing a lot of information. Naively, we could think that we are only betting tiny amounts of money since \$0.25 is a tiny number. However, imagine that we make our example more extreme. In this new game, if the coin lands on heads, you have to pay \$10,000. And if it lands on tails, you get \$10,000.50. The expectation value is again  $\langle x \rangle = \$0.25$ . But the game is much more extreme now. I wouldn't like to play it. While the odds are still in your favor, there is also a real chance that you lose *a lot* of money. This kind of information is encoded in the standard deviation. For this modified game, the standard deviation is  $\Delta x \approx \$10,000.25$ , making it immediately clear that a large sum of money is at stake.

Before we move on, here's one more comment on the idea behind the definition of the standard deviation: The logic behind squaring each possible outcome in the first term on the right-hand side is to avoid that terms cancel because they have op-

posite signs. This is exactly what happens when we calculate the expectation value. In the coin example, if the outcome was heads we had to pay \$1 which mathematically means  $x_1 = -\$1$ . Thus, in the formula for the expectation value, the two terms canceled almost completely since they have different signs. But now we want to get information about the spread in our measurements. Hence, we need to define a notion that takes the *absolute distance* from the expectation value into account. We accomplish this by squaring each outcome and only then multiplying each term by the corresponding probability. We then also square the expectation value such that it has the same units as the thing we just calculated. Otherwise, we would compare apples with pears<sup>8</sup>.

<sup>8</sup>In the example  $\$^2$  with \$.

Now, we have everything we need to get back to quantum mechanics.

## 3.2 Statistical Tools in Quantum Mechanics

After this short detour, let's recall what we have already learned: We describe our system with an abstract object  $|\Psi\rangle$  called a ket. If we want to know, for example, the momentum of the system, we act with the momentum operator  $\hat{p}$  on this object. What we get this way is the value that we would measure for the momentum. In general, however, we are dealing with inherently uncertain quantum systems. Hence, we usually don't get back a simple number if we act on a ket with the momentum operator. For example, consider the ket in Eq. 3.4:

$$|\Psi\rangle = a |\Psi_1\rangle + b |\Psi_2\rangle . \quad (3.10)$$

For this ket, we have

$$\hat{p} |\Psi\rangle = ap_1 |\Psi_1\rangle + bp_2 |\Psi_2\rangle . \quad (3.11)$$

In physical terms, this means that we sometimes measure  $p_1$  and sometimes we measure  $p_2$ . The situation is therefore analogous to the coin example that we talked about in the previous section. Thus, it makes sense to introduce expectation values in

quantum mechanics. Before we can calculate expectation values, we need to add a few more ingredients to our framework.

The multitude of new concepts and ideas can be overwhelming at first, so don't worry if not everything is clear immediately. It takes some time to get used to new concepts. If you feel overwhelmed, just keep reading. After a while, everything will start to fall into place.

I will simply list all of the new ideas and concepts that we need. Afterward, we will talk about them one after another in more detail. Ready?

1. The two ingredients which we need to calculate an expectation value are a list of possible outcomes plus the probabilities with which they occur. If we look at Eq. 3.11, we see that the possible outcomes for this particular ket are  $p_1$  and  $p_2$ . But what are the corresponding probabilities of measuring these values? For our example ket in Eq. 3.11, the probability of measuring  $p_1$  is  $|a|^2$  and the probability of measuring  $p_2$  is  $|b|^2$ . So the correct formula for the momentum expectation value is

$$\langle p \rangle = |a|^2 p_1 + |b|^2 p_2, \quad (3.12)$$

analogous to what we discussed in the previous section.

2. In quantum mechanics, given a concrete ket  $|\Psi\rangle$ , we calculate the expectation value of an operator like  $\hat{p}$  as follows:

$$\begin{aligned}
 \langle \Psi | \hat{p} | \Psi \rangle &\stackrel{\text{Eq. 3.10}}{=} \left( a^* \langle \Psi_1 | + b^* \langle \Psi_2 | \right) \hat{p} \left( a |\Psi_1\rangle + b |\Psi_2\rangle \right) \\
 &= \left( a^* \langle \Psi_1 | + b^* \langle \Psi_2 | \right) \left( a \hat{p} |\Psi_1\rangle + b \hat{p} |\Psi_2\rangle \right) \\
 &= \left( a^* \langle \Psi_1 | + b^* \langle \Psi_2 | \right) \left( a p_1 |\Psi_1\rangle + b p_2 |\Psi_2\rangle \right) \\
 &= |a|^2 p_1 \langle \Psi_1 | \Psi_1 \rangle + a^* b p_2 \langle \Psi_1 | \Psi_2 \rangle + b^* a p_1 \langle \Psi_2 | \Psi_1 \rangle \\
 &\quad + |b|^2 p_2 \langle \Psi_2 | \Psi_2 \rangle \\
 &= |a|^2 p_1 + |b|^2 p_2. \tag{3.13}
 \end{aligned}$$

↴  $a^* a = |a|^2, b^* b = |b|^2,$   
 ↴  $\langle \Psi_1 | \Psi_2 \rangle = 0, \langle \Psi_1 | \Psi_1 \rangle = 1, \dots$

Here we used lots of new ideas that we will discuss in detail below.<sup>9</sup>

<sup>9</sup> We will talk about the new symbols like  $\langle \Psi_1 |$  and why the coefficients  $a$  and  $b$  appear complex conjugated in a moment.

3. A new type object appears in the definition of the expectation value in Eq. 3.13:  $\langle \Psi |$ . We call this kind of object a **bra**. This name together with the name ket is a wordplay on the **brackets** we use. A bra is nothing completely new. Given a ket, we can immediately compute the corresponding bra:

$$\langle \Psi | \equiv |\Psi\rangle^\dagger = (|\Psi\rangle^*)^T. \quad (3.14)$$

The star  $*$  denotes complex conjugation and the superscript  $T$  denotes transposition. The object  $\dagger$  is called **dagger** and we use it to denote a transformation known as **Hermitian conjugation**. We calculate the Hermitian conjugate of an object by first complex conjugating and then transposing it. So, in words, Eq 3.14 means that a bra is the Hermitian conjugate of the corresponding ket.

4. Additionally, in Eq 3.13 we used the fact that our objects  $\langle \Psi_1 |$  and  $\langle \Psi_2 |$  are **normalized basis vectors**. This means that they are orthogonal and thus  $\langle \Psi_1 | \Psi_2 \rangle = 0$ . Moreover, since they are normalized we have  $\langle \Psi_1 | \Psi_1 \rangle = 1$ . This is what we used to get to the last line in Eq 3.13.

Now let's talk about these ideas one after another in detail. We will start at the bottom of the list.

4. As mentioned above, kets (like  $|\Psi_1\rangle$  and  $|\Psi_2\rangle$ ) that correspond to definite values of a given observable are **basis vectors**. This is in contrast to general kets, which do not correspond to one definite value (here  $p_1$  and  $p_2$ ) of, say, the momentum operator but instead to a superposition of possible outcomes. Above, we have expanded the general ket in terms of momentum basis kets<sup>10</sup>

$$|\Psi\rangle = a |\Psi_1\rangle + b |\Psi_2\rangle. \quad (3.15)$$

This is completely analogous to how we can expand an arbi-

<sup>10</sup> Momentum basis kets are those with definite momentum. A general state is a linear combination of momentum basis kets.

trary vector in terms of the basis vectors

$$\vec{e}_x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{e}_z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.16)$$

For example,

$$\vec{v} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = 1\vec{e}_x + 3\vec{e}_y + 5\vec{e}_z = 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 5 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.17)$$

A useful property of such basis vectors is that they are **orthogonal**<sup>11</sup>. Two vectors are orthogonal when their scalar product is zero. For example,  $\vec{e}_x \cdot \vec{e}_y = 0$  and, in general,  $\vec{e}_i \cdot \vec{e}_j = 0$  for  $i \neq j$ . In addition, they are **normalized**. This means that the scalar product of such a basis vector with itself yields one:<sup>12</sup>  $\vec{e}_i \cdot \vec{e}_i = 1$ .

This is why we used in Eq. 3.13 that  $\langle \Psi_1 | \Psi_2 \rangle = 0$ ,  $\langle \Psi_2 | \Psi_1 \rangle = 0$  and in addition,  $\langle \Psi_1 | \Psi_1 \rangle = 1$ ,  $\langle \Psi_2 | \Psi_2 \rangle = 1$ .

3. Next, we need to talk about the new objects that we introduced above:  $\langle \Psi |$ . We introduce these bras specifically to calculate expectation values and concrete probabilities. Expectation values and probabilities are scalar quantities (i.e., ordinary numbers) and not vectors or matrices. So in other words, bras are tools that we add to our framework in order to combine them with our kets  $|\Psi\rangle$  and get back a scalar. The whole point of bras is that if we combine them with a ket, we get back a number.

Again, we can better understand this idea by considering the analogous situation for more familiar, ordinary vectors. For each such vector, we can define a "dual" vector such that the product of a vector and a dual vector yields a number. In other words, vector and dual vector together yield the scalar product. While a normal vector is a column vector, the corresponding dual vector is a row vector. Then, with the rules of matrix multiplication (row times column), a dual

<sup>11</sup> This is not always the case, but for the moment we will only work with orthogonal basis vectors.

<sup>12</sup> Again, this is not necessarily the case. But here, we restrict ourselves to normalized basis vectors.

vector and a normal vector together really yield a number (and not another vector or a matrix):

$$\vec{v}^T \vec{v} = (1, 3, 5) \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = 35.$$

A bra is completely analogous to a row vector while a ket is analogous to a column vector. The idea is that a bra and a ket together yield a number. In other words, this is a convenient way to write the **scalar product**.<sup>13</sup> The transformation that turns a column vector into a row vector is called transposition. Similarly, to get the corresponding bra from a given ket, we transpose it and *additionally* complex conjugate it. This conjugation is necessary since kets are, in general, complex.<sup>14</sup> For complex vectors, we need the additional complex conjugation because we want to interpret the scalar product of a vector with itself as the length of the vector squared:

$$\text{length}^2 = \vec{v} \cdot \vec{v}. \quad (3.18)$$

A complex result wouldn't make sense. The complex conjugation makes sure that we end up with something real. For this reason, the scalar product of *complex* vectors is defined with an additional complex conjugation that makes sure we end up with a real number.<sup>15</sup>

This is also why we defined a bra in Eq. 3.14 as  $\langle \Psi | \equiv |\Psi\rangle^\dagger = (|\Psi\rangle^*)^T$ , and thus why  $a^*$  and  $b^*$  appear in Eq. 3.13.

2. Next, let's take another look at our first calculation of an expectation value in quantum mechanics (Eq. 3.13) which I have rewritten here for convenience:

$$\begin{aligned}
 \langle \Psi | \hat{\rho} | \Psi \rangle &= (a^* \langle \Psi_1 | + b^* \langle \Psi_2 |) (a p_1 |\Psi_1\rangle + b p_2 |\Psi_2\rangle) \\
 &= |a|^2 p_1 \underbrace{\langle \Psi_1 | \Psi_1 \rangle}_{=1} + a^* b p_2 \underbrace{\langle \Psi_1 | \Psi_2 \rangle}_{=0} + b^* a p_1 \underbrace{\langle \Psi_2 | \Psi_1 \rangle}_{=0} \\
 &\quad + |b|^2 p_2 \underbrace{\langle \Psi_2 | \Psi_2 \rangle}_{=1} \\
 &= |a|^2 p_1 + |b|^2 p_2. \tag{3.19}
 \end{aligned}$$

<sup>13</sup> The notion of scalar product means that we multiply two quantities and get back an ordinary number. "Scalar" is a different name for an ordinary number. There are other ways to multiply objects that do not yield an ordinary number like, for example, the cross product. An alternative name for the scalar product is **inner product**.

<sup>14</sup> We will see later why this is the case. For the moment we could say that we simply want to stay general and for this reason, include the complex conjugation which does no harm if the kets are real. But to spoil the surprise: Our kets are complex since kets that describe physical systems are solutions of the Schrödinger equation. And solutions of the Schrödinger equation are, in general, complex.

<sup>15</sup> Because for  $z = a + ib$ , we have  $z^* = a - ib$  and therefore  $z^* z = (a + ib)(a - ib) = a^2 + b^2$  which is real.

The final result has the structure that we talked about in the previous section. We weigh each possible outcome (here  $p_1$  and  $p_2$ ) with the corresponding probability (here  $P(p_1) = |a|^2$  and  $P(p_2) = |b|^2$ ).

1. The last thing we need to talk about is why our probabilities are given by  $P(p_1) = |a|^2$  and  $P(p_2) = |b|^2$ . Naively, we would have guessed that the probabilities of measuring  $p_1$  or  $p_2$  are merely  $a$  and  $b$ , respectively. But kets are, in general, *complex*.<sup>16</sup> This means that coefficients like  $a$  and  $b$  are, in general, complex numbers. Since complex probabilities make no sense, we can't use coefficients like  $a$  and  $b$  directly. Instead, we can use  $|a|^2$  and  $|b|^2$  which are certainly real. So in particular for our example ket in Eq. 3.11, the probability of measuring  $p_1$  is  $|a|^2$  and the probability of measuring  $p_2$  is  $|b|^2$ . The coefficients obtained by expanding a ket in some specific basis, like  $a$  and  $b$  from above, are called **probability amplitudes**.

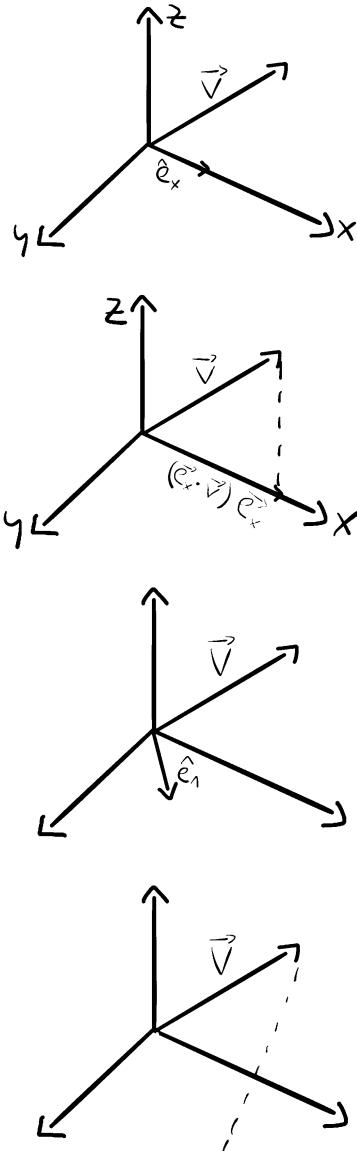
<sup>16</sup> We will see why this is the case in Section 3.6. As already mentioned above, the crucial point is that physical kets are solutions of the Schrödinger equation and solutions of the Schrödinger equation are, in general, complex.

With all this in mind, we can tackle another question typically asked in quantum mechanics: What's the probability of measuring, say, the value  $p_1$ ? For our simple example above, we already know that the probability to measure  $p_1$  is  $|a|^2$ . We are interested, however, in figuring out how to extract this information from any general system. The crucial idea is to use the fact that the state of the system with definite momentum  $p_1$  is  $|\Psi_1\rangle$ . Let's see what happens when we multiply our state  $|\Psi\rangle$  by the bra that corresponds to  $|\Psi_1\rangle$ :

$$\begin{aligned} \langle \Psi_1 | \Psi \rangle &\stackrel{\text{Eq. 3.4}}{=} \langle \Psi_1 | (a |\Psi_1\rangle + b |\Psi_2\rangle) \\ &= a \underbrace{\langle \Psi_1 | \Psi_1 \rangle}_{=1} + b \underbrace{\langle \Psi_1 | \Psi_2 \rangle}_{=0} \\ &= a \times 1 + b \times 0 = a. \end{aligned}$$

We get exactly the result that we wanted:  $a$ . To get the probability, all we have to do is to square the absolute value of this result.

This trick always works. A given ket  $|\Psi\rangle$  possibly consists of dozens or even infinitely many terms when expanded in terms of basis states. But to get the probability of measuring one specific momentum value, we simply have to multiply it by the corresponding conjugated basis state, e.g.,  $\langle\Psi_1|$ .



What we really do here is to project out exactly the term that we want from the sum. Since the basis vectors are orthogonal, the term that survives when multiplied by  $\langle\Psi_1|$  is the one that we are interested in.

This whole procedure is analogous to how we can determine how much a given vector spreads out, say, in the  $z$ -direction. All we have to do is multiply the vector by the  $z$  basis vector  $\vec{e}_z$ :

$$\vec{e}_z \cdot \vec{v} = \vec{e}_z^T \vec{v} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = 5.$$

We can make the analogy even more explicit by using the notation from Eq. 3.17:

$$\begin{aligned} \vec{e}_z \cdot \vec{v} &= \vec{e}_z \cdot (1\vec{e}_x + 3\vec{e}_y + 5\vec{e}_z) \\ &= 1\vec{e}_z \cdot \vec{e}_x + 3\vec{e}_z \cdot \vec{e}_y + 5\vec{e}_z \cdot \vec{e}_z \\ &= 1 \times 0 + 3 \times 0 + 5 \times 1 = 5. \end{aligned} \quad (3.20)$$

### 3.3 Wave Functions

A question you certainly have at this point is: What does all this have to do with the waves we talked about previously? Well, the **wave functions** everyone is talking about are simply the probability amplitudes for one specific basis. Again, let's consider the arbitrary ket  $|\Psi\rangle$  that describes our system. But now, let's say we are interested in possible locations and not the momenta. This means that we have to switch our basis and express  $|\Psi\rangle$  in terms of position basis states.

How can we do this?

The magical tools which help us switch between bases are called **projection operators**. These projection operators aren't unique to quantum mechanics. Projection operators also exist for ordinary vectors. In the previous section, we used the most common basis vectors  $\vec{e}_x, \vec{e}_y, \vec{e}_z$  (Eq. 3.16). However, an equally good (orthogonal and normalized) choice for the basis vectors is:<sup>17</sup>

$$\vec{e}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.21)$$

How can we calculate what our vector  $\vec{v} = (1, 3, 5)^T$  looks like in this basis? We already discussed how we can determine how much a given vector spreads out in any given direction. For example, to find out how much  $\vec{v}$  spreads out in the  $z$ -direction, we multiplied it by  $\vec{e}_z$  (Eq. 3.20). Here, Eq. 3.21 defines new axes relative to the old ones. Therefore, to find out how much  $\vec{v}$  spreads out in the direction defined by  $\vec{e}_1$ , we calculate the scalar product of the two vectors.

$$\tilde{e}_1 \cdot \vec{v} = \frac{1}{\sqrt{2}} (1, 1, 0) \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = \frac{4}{\sqrt{2}}.$$

Analogously, we can calculate how much  $\vec{v}$  spreads out in the other two new directions:

$$\begin{aligned} \tilde{e}_2 \cdot \vec{v} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}^T \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = \frac{-2}{\sqrt{2}}. \\ \tilde{e}_3 \cdot \vec{v} &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}^T \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = 5. \end{aligned}$$

This tells us that, in the new basis (Eq. 3.21), our vector  $\vec{v}$  reads as:

$$\vec{v} = \frac{4}{\sqrt{2}} \vec{e}_1 - \frac{2}{\sqrt{2}} \vec{e}_2 + 5 \vec{e}_3 \hat{=} \begin{pmatrix} \frac{4}{\sqrt{2}} \\ \frac{-2}{\sqrt{2}} \\ 5 \end{pmatrix}_{\text{new basis}}.$$

The general method to rewrite a vector in a new basis is therefore:

<sup>17</sup> The factors  $\frac{1}{\sqrt{2}}$  are normalization constants that make sure that our vectors have length 1.

1. Calculate the scalar product of the vector with each new basis vector.
2. Multiply each result with the corresponding basis vector.
3. The vector in the new basis is the sum of all terms calculated in step 2.

So mathematically, we have

$$\vec{v}_{\text{new basis}} = \sum_i (\vec{e}_i) \underbrace{(\vec{e}_i \cdot \vec{v})}_{\text{a number}} . \quad (3.22)$$

To convince you that this formula is really correct, let's again consider our example from above:

$$\begin{aligned} \vec{v}_{\text{new basis}} &= \sum_i (\vec{e}_i) (\vec{e}_i \cdot \vec{v}) \\ &= (\vec{e}_1)(\vec{e}_1 \cdot \vec{v}) + (\vec{e}_2)(\vec{e}_2 \cdot \vec{v}) + (\vec{e}_3)(\vec{e}_3 \cdot \vec{v}) \\ &= \vec{e}_1 \frac{4}{\sqrt{2}} + \vec{e}_2 \frac{-2}{\sqrt{2}} + 5\vec{e}_3 \quad \checkmark \end{aligned}$$

We use *exactly* the same method in quantum mechanics. We have the ket in the momentum basis,

$$|\Psi\rangle = a |\Psi_1\rangle + b |\Psi_2\rangle , \quad (3.23)$$

and want to calculate how it looks like in the position basis:

$$|\Psi\rangle = c |x_1\rangle + d |x_2\rangle . \quad (3.24)$$

In other words, we want to calculate the coefficients  $c$  and  $d$ . For simplicity, we assume that only two locations,  $x_1$  and  $x_2$ , are possible. In addition, we use a more suggestive notation:  $|x_1\rangle$  is the configuration of the system where we will definitely find our particle at location  $x_1$ , similar to how  $|\Psi_1\rangle$  corresponds to the configuration with momentum  $p_1$ . Using the algorithm we just discussed, we calculate

$$\begin{aligned} |\Psi\rangle &= \sum_i |x_i\rangle \langle x_i| \Psi \rangle \\ &= |x_1\rangle \langle x_1| \Psi \rangle + |x_2\rangle \langle x_2| \Psi \rangle \\ &= c |x_1\rangle + d |x_2\rangle , \end{aligned}$$

where we defined the **probability amplitudes in the position basis**

$$\begin{aligned} c &\equiv \langle x_1 | \Psi \rangle \\ d &\equiv \langle x_2 | \Psi \rangle . \end{aligned}$$

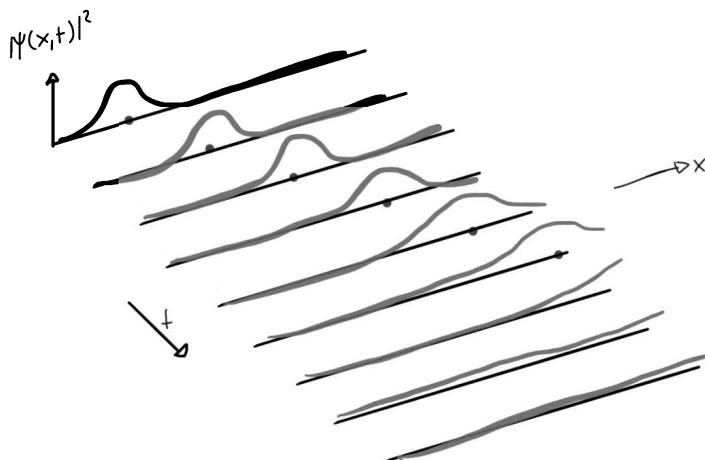
In general, there is a continuum of possible locations and not just a discrete set. Luckily, we can take this into account by simply replacing this sum with an integral:<sup>18</sup>:

$$\begin{aligned} |\Psi\rangle &= \int dx |x\rangle \langle x| \Psi \rangle \\ &= \int dx \Psi(x) |x\rangle , \end{aligned} \quad (3.25)$$

where we defined  $\Psi(x) \equiv \langle x | \Psi \rangle$ . This function  $\Psi(x)$  is analogous to the coefficients we have discussed previously (i.e.,  $a$ ,  $b$ ,  $c$ , and  $d$ ). But we now have one coefficient for *each* location  $x$ .<sup>19</sup> The physical interpretation of  $\Psi(x)$  is again as a probability amplitude. To make the connection to our previous notation explicit, we can write

$$\begin{aligned} \Psi(x_1) &\equiv c \\ \Psi(x_2) &\equiv d . \end{aligned}$$

This probability amplitude  $\Psi(x)$  has a special name: the **wave function**. Completely analogous to what we discussed before, if we square the absolute value of  $\Psi(x)$ , we get the probability to find the particle at a given location.<sup>20</sup>



<sup>18</sup> An integral is, in some sense, the continuum limit of a sum. If we make the steps in a sum smaller and smaller, we end up with an integral.

<sup>19</sup> Take note that  $|\Psi\rangle$  and therefore also  $\Psi(x)$  depend on the time  $t$ . Here we suppress this dependence to unclutter the notation. We will revisit this time dependence later on. Keep in mind, however, that wavefunctions generally depend not only on position, but also on time.

<sup>20</sup> This is not entirely correct. What we get is the probability to find the particle within the interval  $[x, x + dx]$ , since  $\Psi(x)$  is defined under an integral. Thus to get a probability, we not only have to take the absolute square but also integrate over some region. In the image below, the absolute square of a wave function at different points in time is shown. The absolute square indicates the probability of finding the particle in question at different locations. The higher the amplitude of the absolute square of the wave function, the higher the probability. When we are 100% certain where our particle is, the wave function of the particle is the delta distribution (see Appendix C). In some sense, the delta distribution can be thought of as an infinitely thin/localized but also infinitely high function that, however, yields exactly one if we integrate it all over space.

The name *wave* function will make more sense in Section 3.7, in which we will see that functions like  $\Psi(x)$  really exhibit wave-like behavior. One crucial observation is that everything we need to know about the system is encoded in the function  $\Psi(x)$ .

This is analogous to how we usually only write down the coefficients of a vector  $\vec{v} = (1, 3, 5)^T$  without referring to the basis explicitly:  $1\vec{e}_x + 3\vec{e}_y + 5\vec{e}_z$ . Similarly, we use in quantum mechanics  $\Psi(x)$  and don't write  $\int dx \Psi(x) |x\rangle$  all the time.

The trick we used in Eq. 3.25 is generally applicable; it works for any complete basis, not just the position basis. So instead of expanding our ket  $|\Psi\rangle$  in terms of the position basis vectors  $|x\rangle$ , we can also expand it in terms of momentum basis states:

$$\begin{aligned} |\Psi\rangle &= \int dp |p\rangle \langle p|\Psi\rangle \\ &\equiv \int dp \Psi(p) |p\rangle . \end{aligned} \tag{3.26}$$

<sup>21</sup> A general rule that is useful to remember is that bra times ket yields a number (or function):  $\langle x|\Psi\rangle = \Psi(x)$ . And ket times bra yield an operator:  $|x\rangle \langle x|$ . This is analogous to how the usual scalar product of two vectors yields a number:  $\vec{v} \cdot \vec{w} \equiv \vec{v}^T \vec{w} \in \mathbb{R}$ , while  $\vec{v}\vec{w}^T$  yields a matrix. (Remember the general rule for multiplying matrices: "row times column".)

<sup>22</sup> Using that  $\langle\Psi| \equiv |\Psi\rangle^\dagger$  (Eq. 3.14) and

$|\Psi\rangle \equiv \int dx \Psi(x) |x\rangle$   
(Eq. 3.25) we find immediately

$$\langle\Psi| \equiv \int dx \langle x| \Psi^\dagger(x) \tag{3.27}$$

Take note that we could equally write  $\Psi^*(x)$  instead of  $\Psi^\dagger(x)$  since  $\Psi(x)$  is an ordinary function and therefore transposing it makes no difference. Moreover, the symbol  $\delta(x - x')$  is called delta distribution and is for integrals what the Kronecker delta  $\delta_{ij}$  is for sums. For more information, see Appendix C.

The function  $\Psi(p)$  is known as the **momentum representation** of the wave function.

Take note that by comparing the left-hand side with the right-hand side in the first row of Eq. 3.25, we can conclude that the operator  $\int dx |x\rangle \langle x|$  is the identity operator, i.e., an operator that does not change anything.<sup>21</sup>

Analogously, from the first row in Eq. 3.26, we learn that the operator  $\int dp |p\rangle \langle p|$  is also an identity operator:  $\int dp |p\rangle \langle p| \equiv 1$ . This is important since it means that we can always insert  $\int dx |x\rangle \langle x|$  or  $\int dp |p\rangle \langle p|$  in our equations without changing anything. In a moment, we will talk about an example where this is indeed useful.

An important idea is that by using a given wave function  $\Psi(x)$ , we can immediately calculate important quantities like the expectation value:<sup>22</sup>

$$\begin{aligned}
\langle \Psi | \hat{O} | \Psi \rangle &\stackrel{\text{Eq. 3.25}}{=} \langle \Psi | \hat{O} \int dx \Psi(x) | x \rangle \\
&= \int dx' \langle x' | \Psi^\dagger(x') \hat{O} \int dx \Psi(x) | x \rangle && \curvearrowright \text{Eq. 3.27} \\
&= \int dx' \int dx \langle x' | \Psi^\dagger(x') \hat{O} \Psi(x) | x \rangle && \curvearrowright \\
&= \int dx' \int dx \Psi^\dagger(x') \hat{O} \Psi(x) \langle x' | x \rangle && \curvearrowright \\
&= \int dx' \int dx \Psi^\dagger(x') \hat{O} \Psi(x) \delta(x - x') && \curvearrowright \langle x' | x \rangle = \delta(x - x') \\
&= \int dx \Psi^\dagger(x) \hat{O} \Psi(x). && \curvearrowright \int dx' f(x') \delta(x - x') = f(x) \quad (3.28)
\end{aligned}$$

To get to the last line, we used that if we act with  $\hat{O}$  on our wave function  $\Psi(x)$ , we get a number and that we can move numbers around freely. In addition, the basic states  $|x\rangle$  are orthogonal and normalized. Therefore, a scalar product like  $\langle x' | x \rangle$  yields zero except when  $x' = x$ . This is analogous to how the scalar product between two basis vectors  $\vec{e}_i \cdot \vec{e}_j$  yields zero, except when  $i = j$ . So when we integrate over  $x'$ , all terms vanish except for where  $x = x'$ . This is what the Dirac delta distribution  $\delta(x - x')$  encodes, similar to how the Kronecker delta  $\delta_{ij}$  picks the element from a given sum where  $i = j$ .<sup>23</sup>

We can also perform the same steps as in Eq. 3.28 without the operator  $\hat{O}$  and with a bra that is different from the ket:

$$\begin{aligned}
\langle \Phi | \Psi \rangle &= \int dx' \langle x' | \Phi^\dagger(x') \int dx \Psi(x) | x \rangle && \text{using Eq. 3.25 and Eq. 3.27} \\
&= \int dx' \int dx \langle x' | \Phi^\dagger(x') \Psi(x) | x \rangle && \curvearrowright \\
&= \int dx' \int dx \Phi^\dagger(x') \Psi(x) \langle x' | x \rangle && \curvearrowright \\
&= \int dx' \int dx \Phi^\dagger(x') \Psi(x) \delta(x - x') && \curvearrowright \langle x' | x \rangle = \delta(x - x') \\
&= \int dx \Phi^\dagger(x) \Psi(x). && \curvearrowright \int dx' f(x') \delta(x - x') = f(x) \quad (3.29)
\end{aligned}$$

<sup>23</sup> The Kronecker delta and the Dirac delta distribution are discussed in Appendix C.

This is how the scalar product of two kets looks if we use an explicit basis. Note that this is completely analogous to the

scalar product of two complex vectors in an explicit basis:

$$\begin{aligned}
 \vec{w}^* \cdot \vec{z} &= \sum_i \sum_j (w_i^* \vec{e}_i) \cdot (z_j \vec{e}_j) \\
 &= \sum_i \sum_j w_i^* z_j \vec{e}_i \cdot \vec{e}_j \\
 &= \sum_i \sum_j w_i^* z_j \delta_{ij} \\
 &\quad \Rightarrow \vec{e}_i \cdot \vec{e}_j = \delta_{ij} \\
 &= \sum_i w_i^* z_i. \\
 &\quad \Rightarrow \sum_j f_j \delta_{ij} = f_i
 \end{aligned}$$

Don't worry if not everything is perfectly clear at this point. Everything we have discussed here will make a lot more sense as soon as we discuss explicit examples. Our goal is to get an overview and not to understand everything perfectly.

Two crucial questions remain unanswered:

- ▷ What do these quantum operators we've been discussing *really* look like?
- ▷ How do quantum systems evolve as time passes?

This is what we will talk about next.

## 3.4 Quantum Operators

So far, we have talked about quantum operators in quite abstract terms. We simply introduced a symbol  $\hat{p}$  and proposed that it does the job:

$$\hat{p} |\Psi_1\rangle = p_1 |\Psi_1\rangle. \quad (3.30)$$

Luckily, there is an important concept that will allow us to move beyond these abstractions: Noether's theorem.<sup>24</sup>

Discussing the theorem in full detail would lead us too far astray. But all we need is the punchline:

<sup>24</sup> There are, in fact, two famous theorems by Emmy Noether that are relevant in this context. The one I'm talking about is her first one that deals with *global* symmetries. Her second one is about *local* symmetries.

**Symmetries lead to conserved quantities.**

The most important examples:

- ▷ If the system does not change under rotations, we know immediately that angular momentum is conserved. In other words, if we can rotate our system without changing anything, then angular momentum is conserved.
- ▷ If the system does not change under spatial translations  $x \rightarrow x + \epsilon$ , we know immediately that momentum is conserved. This means that if we change the position of the whole system and nothing changes, then momentum is conserved.
- ▷ If the system does not change under temporal translations  $t \rightarrow t + \epsilon$ , we know immediately that energy is conserved. Formulated differently, if the system behaved yesterday exactly as it does today, energy is conserved.

This is one puzzle piece. The second piece we need is how symmetries are described mathematically.

You are probably wondering what all this has to do with quantum mechanics. As we will see after the next section, actually, a lot! The mathematical ideas in the next section are exactly what we need to find the explicit form of quantum operators. Most importantly, this explicit form allows us to derive the famous canonical commutation relation, and the Schrödinger equation.

So let's talk about symmetries in mathematical terms.

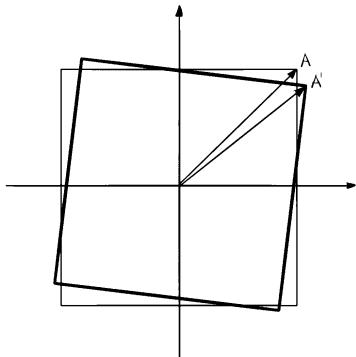
### 3.4.1 Intermezzo: Symmetries

The role and description of symmetries in physics is a huge topic.<sup>25</sup> So there is no way we can cover all the details. However, I will try to emphasize the basic ideas that are necessary to

<sup>25</sup> In fact, I've written a whole book on exactly this topic:

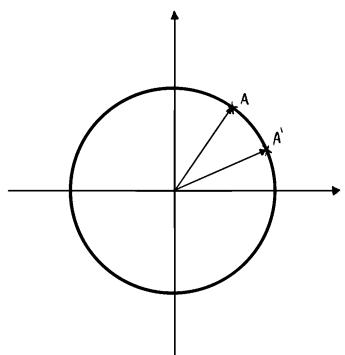
Jakob Schwichtenberg. *Physics from Symmetry*. Springer, Cham, Switzerland, 2018. ISBN 978-3319666303

understand quantum mechanics.



First of all: what is a symmetry?

Imagine a friend stands in front of you and holds an object in her hands. Then you close your eyes and she performs a transformation of the object (e.g. a rotation). Then you open your eyes again. If you can't tell if your friend changed anything at all, the transformation she performed is a symmetry of the object. For example, if she holds a perfectly round, single-colored ball in her hands, any rotation is a symmetry of the ball. In contrast, if she holds a box in her hand, only very specific rotations are symmetries of the box. Doing nothing is always a symmetry.



The bottom line is:

**A symmetry is a transformation that leaves the object in question unchanged.**

The part of mathematics which deals with symmetries is called **group theory**. A group is a set of transformations which fulfill special rules plus an operation that tells us how to combine the transformations. The rules are known as group axioms and we can motivate them by investigating an intuitive symmetry like rotational symmetry. We will not discuss details like this since we don't need them for what follows.

Also, we only need one special part of group theory, namely the part that deals with *continuous* symmetries. An example of a continuous symmetry is the one I just mentioned: rotations of a ball. Rotations are continuous because we can label them with a continuous parameter: the angle of rotation. In contrast, there are also *discrete* symmetries. The most famous examples are mirror symmetries.

There is one property that makes continuous symmetries especially nice to deal with: they have elements which are arbitrarily close to the identity transformation.<sup>26</sup>

<sup>26</sup> The identity transformation is the transformation that changes nothing at all. For example, for rotations, the identity transformation is a rotation by  $0^\circ$ .

For example, think about the symmetries of a circle. Any rotation about the origin is a symmetry of a circle. Therefore, a rotation extremely close to the identity transformation, say a rotation by  $0.000001^\circ$ , is a symmetry of the circle.

In contrast, an arbitrary group has, in general, no element close to the identity.

For example, think about the symmetries of a square. The set of transformations that leaves a square invariant comprises four rotations: a rotation by  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  and  $270^\circ$ , plus some mirror symmetries. But a rotation by  $0.000001^\circ$  is not a symmetry.

Mathematically, we write an element  $g$  close to the identity  $I$  as:

$$g(\epsilon) = I + \epsilon G, \quad (3.31)$$

where  $\epsilon$  is a really, really small number and  $G$  is an object, called a **generator**, that we will talk about in a moment. In the smallest possible case, such transformations are called **infinitesimal transformations**.

Such small transformations barely change anything. However, if we repeat an infinitesimal transformation many times, we end up with a finite transformation.

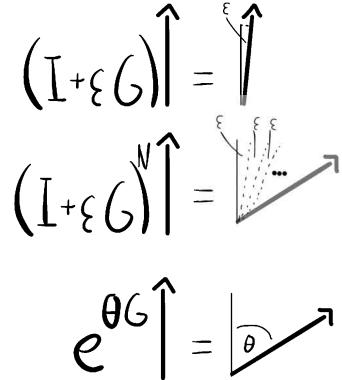
Let's return to our discussion about rotations. Many small rotations in one direction are equivalent to one big rotation in the same direction.

Mathematically, we can write the idea of repeating a small transformation many times as follows

$$h(\theta) = (I + \epsilon G)(I + \epsilon G)(I + \epsilon G)\dots = (I + \epsilon G)^N, \quad (3.32)$$

where  $N$  denotes how often we repeat the small transformation.

If  $\theta$  denotes some finite transformation parameter, e.g.,  $50^\circ$  or so, and  $N$  is some huge number that makes sure we are close to



the identity, we can write Eq. 3.31 as

$$g(\theta) = I + \frac{\theta}{N} G. \quad (3.33)$$

The transformations we want to consider are the smallest possible, which means  $N$  must be the biggest possible number, i.e.,  $N \rightarrow \infty$ . To get a finite transformation from such an infinitesimal transformation, one has to repeat the infinitesimal transformation infinitely often. Mathematically we can write this as

$$h(\theta) = \lim_{N \rightarrow \infty} \left( I + \frac{\theta}{N} G \right)^N, \quad (3.34)$$

<sup>27</sup> This is one possible definition of the exponential function. We derive another definition in terms of an infinite series in Appendix A.

which is in the limit  $N \rightarrow \infty$  the exponential function<sup>27</sup>

$$h(\theta) = \lim_{N \rightarrow \infty} \left( I + \frac{\theta}{N} G \right)^N = e^{\theta G}. \quad (3.35)$$

The bottom line is that the object  $G$  *generates* the finite transformation  $h$ . This is why we call objects like this **generators**.

What do these generators explicitly look like?

Let's consider a function  $f(x, t)$  and assume that our goal is to generate a spatial translation such that  $Tf(x, t) = f(x + a, t)$ . The generator

$$G_{\text{trans}} = \partial_x \quad (3.36)$$

<sup>28</sup> Here  $\partial_x$  is a shorthand notation for the derivative  $\frac{\partial}{\partial x}$ .

does the job:<sup>28</sup>

$$\begin{aligned} e^{aG_{\text{trans}}} f(x, t) &= \left( 1 + aG_{\text{trans}} + \frac{a^2}{2} G_{\text{trans}}^2 + \dots \right) f(x, t) \\ &= \left( 1 + a\partial_x + \frac{a^2}{2} \partial_x^2 + \dots \right) f(x, t) \\ &= f(x + a, t) \end{aligned}$$

Here we used the series expansion of  $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ . In the last step, we used that in the second to last line we have exactly the Taylor expansion of  $f(x + a, t)$ .<sup>29</sup> Alternatively, consider an

<sup>29</sup> If you're unfamiliar with the Taylor series expansion, have a look at Appendix A.

infinitesimal translation:  $a \rightarrow \epsilon$  with  $\epsilon \ll 1$ . We then have

$$\begin{aligned}
 e^{\epsilon G_{\text{xtrans}}} f(x, t) &= (1 + \epsilon G_{\text{xtrans}} + \frac{\epsilon^2}{2} G_{\text{xtrans}}^2 + \dots) f(x, t) \\
 &\approx (1 + \epsilon G_{\text{xtrans}}) f(x, t) \\
 &= (1 + \epsilon \partial_x) f(x, t) \\
 &= f(x, t) + \epsilon \partial_x f(x, t) = f(x + \epsilon, t). \tag{3.37}
 \end{aligned}$$

$\supset \epsilon^2 \approx 0, \epsilon^3 \approx 0, \dots$   
 $\supset \text{Eq. 3.36}$   
 $\supset (3.37)$

Here  $\partial_x f(x, t)$  is the **rate of change** of  $f(x, t)$  in the  $x$ -direction.

If we multiply this rate of change by the distance  $\epsilon$  that we move in the  $x$ -direction, we end up with the total change of  $f(x, t)$  if we move by  $\epsilon$  in the  $x$ -direction. Thus,  $f(x, t) + \epsilon \partial_x f(x, t)$  really *is* the value of  $f$  at the location  $x + \epsilon$ .

The bottom line is:  $G_{\text{xtrans}} = \partial_x$  generates *spatial* translations.

Completely analogously,  $G_{\text{ttrans}} = \partial_t$  generates *temporal* translations:

$$f(x, t) \rightarrow e^{a G_{\text{ttrans}}} f(x, t) = f(x, t + a).$$

What we have learned here is that generators are the crucial mathematical ingredient that we need to describe continuous symmetries. We can describe *any* continuous symmetry by acting many, many times with the corresponding generator on the function in question.

So we can summarize:

The core of each continuous symmetry  
is the corresponding generator.

This idea is already everything we need to determine *explicitly* what the most important quantum operators look like.

### 3.4.2 The Most Important Quantum Operators

We now put the puzzle pieces together. The pieces we have so far are:

- ▷ We are looking for quantum operators. When we act with, for example, the momentum operator on the ket  $|\Psi_1\rangle$  that describes our system, we want to get  $\hat{p}|\Psi_1\rangle = p_1|\Psi_1\rangle$ .
- ▷ Noether's theorem tells us: there is a deep connection between symmetries and the most important physical quantities: momentum, energy, etc.
- ▷ Group theory tells us that the essential objects which are responsible for these symmetries are the corresponding generators.

The crucial idea is to take Noether's theorem seriously. Instead of saying that we get a conserved quantity if there is a symmetry, we say the operator responsible for the symmetry (the generator) also describes the conserved quantity (the quantum operator).

In slogan form:

**quantum operator  $\leftrightarrow$  generator of symmetry**

This may seem like quite a stretch, but we will see that this idea works incredibly well in the next section.

Now, let's make the above statement explicit.

Momentum is connected to symmetry under spatial translations. Therefore, we make the identification<sup>30</sup>

**momentum  $\hat{p}_i \leftrightarrow$  generator of spatial translations  $(-i\hbar\partial_i)$**

(3.38)

Analogously, energy is connected to symmetry under temporal translations. Therefore

**energy  $\hat{E} \leftrightarrow$  generator of temporal translations  $(i\hbar\partial_t)$**

(3.39)

Take note that we have introduced a new fundamental constant:  $\hbar$ .<sup>31</sup> This constant is known as the **Planck constant** and encodes

<sup>30</sup> This constant already appeared in the formulas for the relationship between the wavelength and the momentum and energy (Eq. 2.2 and Eq. 2.3).

the magnitude of quantum effects.<sup>32</sup> We can understand the need for a new constant by observing that momentum has dimensions  $[p] = \text{kg} \cdot \text{m/s}$ . However,  $\partial_i$  has the unit<sup>33</sup>  $[\partial_i] = 1/\text{m}$ . Similarly,  $\partial_t$  has the unit  $[\partial_t] = 1/\text{s}$ , while energy has the unit  $[E] = \text{kg} \cdot \text{m}^2/\text{s}^2$ . Therefore, we need something to get the same units on the right-hand and left-hand side of the equations. Using the units of energy, momentum and the differential operators we can conclude that  $[\hbar] = \text{kg} \cdot \text{m}^2/\text{s}$  since<sup>34</sup>

$$\begin{aligned}[p] &= \text{kg} \cdot \frac{\text{m}}{\text{s}} \stackrel{!}{=} [-i\hbar\partial_i] = \text{kg} \cdot \frac{\text{m}^2}{\text{s}} \frac{1}{\text{m}} = \text{kg} \cdot \frac{\text{m}}{\text{s}} \quad \checkmark \\ [E] &= \text{kg} \cdot \frac{\text{m}^2}{\text{s}^2} \stackrel{!}{=} [i\hbar\partial_t] = \text{kg} \cdot \frac{\text{m}^2}{\text{s}} \frac{1}{\text{s}} = \text{kg} \cdot \frac{\text{m}^2}{\text{s}^2} \quad \checkmark\end{aligned}$$

The Planck constant is one of the most important fundamental constants and we need to extract its value from experiments:  $\hbar \approx 1.055 \cdot 10^{-34} \text{ kg} \cdot \text{m}^2/\text{s}$ . Since there is no symmetry connected to the conservation of position, the position operator stays as is<sup>35</sup>  $\hat{x}$ . In addition, using the definition of angular momentum ( $\vec{L} = \vec{x} \times \vec{p}$ ), we can, in principle, write down the angular momentum operator by simply replacing  $\vec{p}$  with the corresponding quantum operator  $\hat{p}$ . However, there is an important subtlety that this approach neglects. As such, we will postpone further discussion of this fact until Section 3.10.

Now, equipped with the ideas discussed in this section, we are able to derive one of the most important equations in *all* of physics.

<sup>32</sup> To be precise  $\hbar$  is the reduced Planck constant (speak: "h-bar") defined by  $\frac{h}{2\pi}$  where  $h$  is the real Planck constant. The additional factor  $2\pi$  makes sure that we really get the momentum and energy, as we will see in Section 3.7.

<sup>33</sup> Recall  $\partial_x = \frac{\partial}{\partial x}$ . The symbol  $\partial x$ , in some sense, simply means a tiny amount of  $x$  and therefore has the same unit as  $x$ . Therefore,  $\partial_x \propto 1/\partial x$  has the unit  $1/m$ .

<sup>34</sup> Square brackets around a quantity means that we are talking about the units of the quantity.

<sup>35</sup> This may seem confusing at first. But all this operator does if we act with it on a function  $f(x)$  is to multiply it by  $x$ , i.e.,  $\hat{x}f(x) = xf(x)$  since  $x$  really is the location we evaluate  $f(x)$  at.

### 3.5 How Operators Influence Each Other

As already emphasized above, the crucial new thing in quantum mechanics is the inherent uncertainty. This uncertainty comes about since the momentum changes every time we measure position and the position changes every time we measure momentum. Mathematically, this statement means that it makes a difference whether we first measure the momentum or first measure the location:  $\hat{x}\hat{p}|\Psi\rangle \neq \hat{p}\hat{x}|\Psi\rangle$ . We can also write this

statement as  $(\hat{x}\hat{p} - \hat{p}\hat{x}) |\Psi\rangle \neq 0$ . The shorthand notation for this expression is

$$[A, B] \equiv AB - BA. \quad (3.40)$$

We usually call this the **commutator** of  $A$  and  $B$ .

To get an intuitive understanding for what it means for operators to have a non-zero commutator, compare the situation where you first put on your socks and then your shoes with the situation where you first put on your shoes and then your socks. The outcome is clearly different — the ordering of the operations "putting shoes on" and "putting socks on" therefore matters. In technical terms, we say these two operations do not commute. In contrast, it makes no difference if you put on your left sock first and then your right sock or your right sock first and then your left sock. These two operations do commute.

Now, using the explicit quantum operator  $\hat{p}_i = -i\hbar\partial_i$  (Eq. 3.38), we can actually *derive* that  $[\hat{p}_i, \hat{x}_j] \neq 0$ :

$$\begin{aligned} [\hat{p}_i, \hat{x}_j] |\Psi\rangle &= (\hat{p}_i \hat{x}_j - \hat{x}_j \hat{p}_i) |\Psi\rangle \\ &= (-i\hbar\partial_i \hat{x}_j + \hat{x}_j i\hbar\partial_i) |\Psi\rangle \quad \text{→ Eq. 3.38} \\ &= -(i\hbar\partial_i \hat{x}_j) |\Psi\rangle - \cancel{\hat{x}_j (i\hbar\partial_i |\Psi\rangle)} + \cancel{\hat{x}_j i\hbar\partial_i |\Psi\rangle} \quad \text{→ product rule} \\ &= -i\hbar\delta_{ij} |\Psi\rangle. \quad \text{→ } \partial_i \hat{x}_j = \delta_{ij} \end{aligned} \quad (3.41)$$

<sup>36</sup> The Kronecker delta  $\delta_{ij}$  is zero for  $i \neq j$  and one for  $i = j$ . If you're unfamiliar with it, have a look at Appendix C.

In the last step, we used that, for example,  $\partial_y x = 0$  but  $\partial_y y = 1$ .<sup>36</sup> We didn't assume anything about  $|\Psi\rangle$ , so the equation is valid for any  $|\Psi\rangle$ . Therefore, we can write the equation without it, which makes the equation a bit shorter. However, we always have to remember that there is an implicit ket in such equations and we are only too lazy to write it all the time. In conclusion:

$$[\hat{p}_i, \hat{x}_j] = -i\hbar\delta_{ij} \quad (3.42)$$

This little equation is known as the **canonical commutation relation**. As the name already indicates this equation is extremely

important. In fact, many textbooks and lectures use it as a *starting point* for quantum mechanics.<sup>37</sup>

The canonical commutation relation encodes an incredibly important property of quantum mechanics: It makes a difference in which order we measure the position and the momentum! This is exactly the quantum uncertainty we talked about previously.<sup>38</sup> In physical terms, it means that we can't measure the position and the momentum of a particle at the same time with arbitrary precision. We can now also understand how this comes about in our framework. We identified the momentum operator as the generator of spatial translations (Eq. 3.38). Thus, each time we measure the momentum, we perform a tiny spatial translation. Therefore, what we've discovered here is exactly what we discussed already in Section 2.3, namely that each time we measure the location of an electron, we have to push it a little bit with a photon.

---

The next cool thing we can do using the explicit quantum operators is to derive how quantum systems change in time. This is described by the second most important equation of quantum mechanics, which we will discuss in the next section.

## 3.6 How Quantum Systems Change

So far, we have treated quantum systems as if they were static. In the real world, however, systems generally change over time. Therefore, we need something that allows us to understand how a quantum system evolves as time passes. Mathematically, what we need is an equation of this form:

$$\partial_t |\Psi\rangle = \dots$$

Here,  $\partial_t |\Psi\rangle$  is the rate of change of  $|\Psi\rangle$  if we move with respect to  $t$ . In other words,  $\partial_t |\Psi\rangle$  describes how  $|\Psi\rangle$  changes as time

<sup>37</sup> Formulated differently, many textbooks use this equation as the fundamental postulate of quantum mechanics. For example, using this equation, we can derive what the momentum operator looks like etc.

<sup>38</sup> We talked about the quantum uncertainty in Section 2.3. The connection between the commutation relation and uncertainty can be made mathematically precise. Using mathematical results such as the Cauchy-Schwartz inequality and the explicit definition of the standard deviation it is possible to derive for two general operators  $\hat{A}$  and  $\hat{B}$  the following equation

$$(\Delta\hat{A})^2(\Delta\hat{B})^2 \geq |\frac{1}{2i}\langle[\hat{A}, \hat{B}]\rangle|,$$

where  $\Delta\hat{A}$  and  $\Delta\hat{B}$  are the standard deviations of the two operators. Therefore, using Eq. 3.42 we find

$$(\Delta\hat{p})^2(\Delta\hat{x})^2 \geq |\frac{1}{2i}\langle[\hat{p}, \hat{x}]\rangle| = \frac{\hbar}{2}.$$

A measurement without any variance would mean  $\Delta\hat{p} = 0$  plus  $\Delta\hat{x} = 0$ . This equation tells us that this is impossible. It's impossible to measure the position and the momentum of a particle at the same time with arbitrary precision. If we reduce, for example, the variance in our position measurements, we automatically get a larger variance in our momentum measurements.

passes — exactly what we need! On the right-hand side, we need something that contains specific details about the system in question. Thus, the right-hand side will be different for each system. Then, as soon as we have such an equation, we have to solve it to understand how  $|\Psi\rangle$  depends on  $t$ .

Luckily, we don't have to guess. All we have to do is take a second look at what we derived in the previous section. The main actor here,  $\partial_t$ , is almost exactly the quantum energy operator  $i\hbar\partial_t$  (Eq. 3.39), only without the imaginary unit and Planck's constant. Therefore, we know that we have the energy on the right-hand side of the equation:

$$i\hbar\partial_t |\Psi\rangle = E |\Psi\rangle .$$

Is there anything else we know about energy?

In classical mechanics, the total energy is given as the sum of kinetic energy and potential energy:

$$E = T + V \equiv \text{kinetic energy} + \text{potential energy}. \quad (3.43)$$

The usual formula for the kinetic energy is

$$T = \frac{1}{2}mv^2 = \frac{p^2}{2m} \quad (3.44)$$

since  $p = mv$ . We can turn this equation into a quantum equation by replacing the classical momentum,  $p$ , with the quantum momentum operator,  $\hat{p}_i = -i\hbar\partial_i$ , which we derived above.<sup>39</sup>

The potential energy usually only depends on the position:  $V = V(x)$ . Therefore, we can turn it into a quantum operator by letting  $x \rightarrow \hat{x}$ .<sup>40</sup>

Returning to our equation,

$$i\hbar\partial_t |\Psi\rangle = E |\Psi\rangle ,$$

we can use the usual formula  $E = T + V$  and replace the classi-

<sup>39</sup> The index  $i$  is necessary because, in general, we are dealing with systems that can change in 3-spatial dimensions. The index takes on the values  $i = \{x, y, z\}$  or equivalently  $i = \{1, 2, 3\}$ .

<sup>40</sup> If there is more than one dimension, we have  $\hat{x}_i$ .

cal variables with the corresponding quantum operators:

$$\begin{aligned}
 i\hbar\partial_t |\Psi\rangle &= E |\Psi\rangle \\
 &= (T + V) |\Psi\rangle && \text{Eq. 3.43} \\
 &= \left( \frac{p^2}{2m} + V(x) \right) |\Psi\rangle && \text{Eq. 3.44} \\
 \Rightarrow i\hbar\partial_t |\Psi\rangle &= \left( \frac{\hat{p}^2}{2m} + V(\hat{x}) \right) |\Psi\rangle && \text{introducing operators} \\
 &= \left( \frac{(-i\hbar\partial_i)^2}{2m} + V(\hat{x}) \right) |\Psi\rangle && \text{Eq. 3.38} \\
 &= \left( -\frac{\hbar^2\partial_i^2}{2m} + V(\hat{x}) \right) |\Psi\rangle .
 \end{aligned}$$

The resulting equation,

$$i\hbar\partial_t |\Psi\rangle = -\frac{\hbar^2\partial_i^2}{2m} |\Psi\rangle + V(\hat{x}) |\Psi\rangle , \quad (3.45)$$

is the famous **Schrödinger equation**. Our main job in quantum mechanics will often be to solve the Schrödinger equation for a given potential  $V(x)$  and specific boundary conditions.

We will discuss the most important quantum systems in Chapter 5. Note that for historical reasons, the energy operator on the right-hand side of the Schrödinger equation is known as the **Hamiltonian operator**,

$$\hat{H} \equiv \left( -\frac{\hbar^2\partial_i^2}{2m} + V(\hat{x}) \right) . \quad (3.46)$$

The more general form of the Schrödinger equation,

$$i\hbar\partial_t |\Psi\rangle = \hat{H} |\Psi\rangle \quad (3.47)$$

is valid even for relativistic systems and quantum field theories.

All we have to do is modify  $\hat{H}$  appropriately.<sup>41</sup>

<sup>41</sup> In fact, we can simply replace the classical energy-momentum relation ( $E = \frac{p^2}{2m}$ ) with the relativistic version ( $E^2 = (pc)^2 + (mc^2)^2$ ). Then again we replace  $p$  with the corresponding quantum operator and end up with an equation that describes how relativistic quantum systems evolve in time. The resulting equation is known as Klein-Gordon equation and is the correct equation of motion for (spinless) relativistic particles.

A convenient alternative way to describe the time-evolution of quantum systems is with the so-called **time evolution operator**  $U(t)$ . We define this operator through the following formula

$$|\Psi(x, t)\rangle = U(t)|\Psi(x, 0)\rangle. \quad (3.48)$$

In words, this means that if we act with this operator on some ket  $|\Psi(x, 0)\rangle$ , the resulting ket describes the system at time  $t$ :  $|\Psi(x, t)\rangle$ . This operator is not merely an abstract thing. We can write it down explicitly by putting Eq. 3.48 into our Schrödinger equation:

$$\begin{aligned} i\hbar\partial_t|\Psi(x, t)\rangle &= H|\Psi(x, t)\rangle \\ i\hbar\partial_t U(t)|\Psi(x, 0)\rangle &= HU(t)|\Psi(x, 0)\rangle. \end{aligned} \quad \text{Eq. 3.48}$$

This equation holds for any  $|\Psi(x, 0)\rangle$  and we can therefore write it without it:

$$\begin{aligned} i\hbar\partial_t U(t) &= HU(t) \\ i\hbar \frac{\partial_t U(t)}{U(t)} &= H. \end{aligned}$$

This is a differential equation for  $U(t)$  and the general solution is

$$U(t) = e^{-\frac{i}{\hbar} \int_0^t dt' H(t')} \quad (3.49)$$

since

$$\begin{aligned} i\hbar \frac{\partial_t U(t)}{U(t)} &= H \\ i\hbar \frac{\partial_t e^{-\frac{i}{\hbar} \int_0^t dt' H(t')}}{e^{-\frac{i}{\hbar} \int_0^t dt' H(t')}} &= H \\ i\hbar \left( -\frac{i}{\hbar} \underbrace{\partial_t \int_0^t dt' H(t')}_{=H} \right) \frac{e^{-\frac{i}{\hbar} \int_0^t dt' H(t')}}{e^{-\frac{i}{\hbar} \int_0^t dt' H(t')}} &= H \\ H = H &\checkmark \end{aligned} \quad \text{Eq. 3.49}$$

So the time-evolution operator is simply a convenient way to write the information encoded in the Schrödinger equation a bit differently.

We can now answer one additional question: Where in the quantum framework can we see that quantum mechanics is about waves?

Actually, by looking at the Schrödinger equation! The crucial point is that the Schrödinger equation is a **wave equation**. This means that its solutions behave like waves. We will see this explicitly for the simplest kind of system in the next section.

### 3.7 Why Quantum Mechanics is About Waves

Let's consider a quantum system without any potential ( $V(\hat{x}) = 0$ ). The time evolution of the system is described by the equation

$$\text{Eq. 3.45 with } V = 0 : \quad i\hbar\partial_t |\Psi\rangle = -\frac{\hbar^2\partial_x^2}{2m} |\Psi\rangle .$$

Before we solve this equation, we should choose a specific basis to make things simpler. For this example, we will use the position and therefore the wave functions that we introduced in Section 3.3. Also, to unclutter the notation, we will restrict ourselves to one spatial dimension. We are then left with

$$\begin{aligned} i\hbar\partial_t |\Psi\rangle &= -\frac{\hbar^2\partial_x^2}{2m} |\Psi\rangle \\ i\hbar\partial_t \int dx \psi(x, t) |x\rangle &= -\frac{\hbar^2\partial_x^2}{2m} \int dx \psi(x, t) |x\rangle \quad \text{Eq. 3.25} \\ i\hbar\partial_t \psi(x, t) &= -\frac{\hbar^2\partial_x^2}{2m} \psi(x, t) . \end{aligned}$$

This is a legitimate approach, as long as we remember that  $\psi(x, t)$  is defined under an integral. Hence, it describes a **density**. Since the coefficients in front of the kets are probability amplitudes, the wave function  $\psi(x, t)$  is a **probability amplitude density**. To get probabilities<sup>42</sup>, we have to take the absolute square and then integrate it over some spatial region. The result

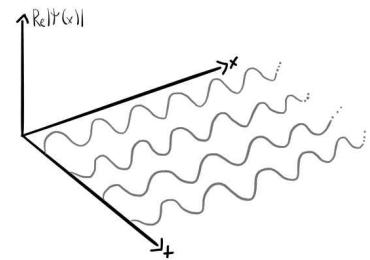


Figure 3.1: Time evolution of a plane wave. Since, in general, a wave function is complex and therefore hard to visualize, only the real part is shown.

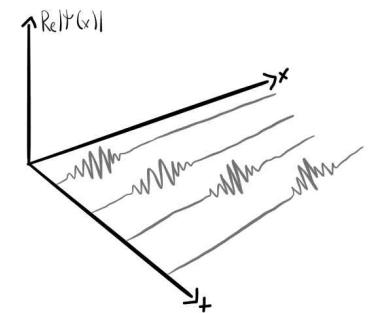


Figure 3.2: Time evolution of a wave packet. Again, only the real part is shown.

<sup>42</sup> After all, probabilities are what we really can measure in experiments.

we obtain is the probability to find the particle in the region that we integrated over. One solution to this equation is

$$\psi(x, t) = e^{-i(Et - px)/\hbar} \quad (3.50)$$

since

$$\begin{aligned} i\hbar \partial_t \psi(x, t) &= -\frac{\hbar^2 \partial_x^2}{2m} \psi(x, t) && \text{Eq. 3.50} \\ i\hbar \partial_t e^{-i(Et - px)/\hbar} &= -\frac{\hbar^2 \partial_x^2}{2m} e^{-i(Et - px)/\hbar} && \partial_x e^{ipx} = ip e^{ipx} \\ E e^{-i(Et - px)/\hbar} &= \frac{p^2}{2m} e^{-i(Et - px)/\hbar} && E = \frac{p^2}{2m} \\ \frac{p^2}{2m} e^{-i(Et - px)/\hbar} &= \frac{p^2}{2m} e^{-i(Et - px)/\hbar} \quad \checkmark, && \end{aligned} \quad (3.51)$$

<sup>43</sup> If you wonder why this is the case, remember Euler's formula:  $e^{ix} = \cos(x) + i \sin(x)$ . The well-known  $\cos(x)$  and  $\sin(x)$  are oscillating functions that perfectly describe waves. In other words, the real and imaginary parts of a function proportional to  $e^{ix}$  describe an oscillation that spreads out all over space with equal amplitude. (Hence the prefix *plane* wave.)

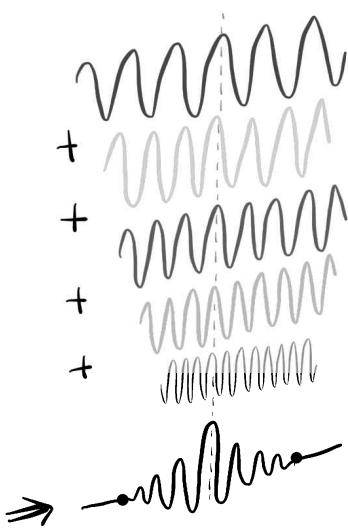


Figure 3.3: Construction of a wave packet as a superposition of plane waves.

where in the last step we used the fact that  $E$  is just the numerical value for the total energy of a free particle:  $E = \frac{p^2}{2m}$ . A function of the form  $e^{-i(Et - \vec{p}\vec{x})/\hbar}$  is known as a **plane wave**.<sup>43</sup>

An important observation: the Schrödinger equation is linear in  $\psi$ . This means that we don't have any terms like  $\psi^2$ ,  $\psi^3$  etc. A direct consequence of this property is that if we have two or more solutions  $\psi_1, \psi_2, \dots$ , we can immediately write down additional solutions:

$$\psi_{\text{sup}} = a\psi_1 + b\psi_2 + \dots \quad (3.52)$$

This is known as a **superposition**. We can see that a superposition is also a solution by putting it into the Schrödinger equation (Eq. 3.47):

$$\begin{aligned} H\psi_{\text{sup}} &= i\hbar \partial_t \psi_{\text{sup}} && \text{(this is Eq. 3.47)} \\ H(a\psi_1 + b\psi_2 + \dots) &= i\hbar \partial_t(a\psi_1 + b\psi_2 + \dots). && \text{Eq. 3.52} \end{aligned} \quad (3.53)$$

The functions  $\psi_1, \psi_2, \psi_3, \dots$  are solutions, which means

$$H\psi_1 = E_1\psi_1, \quad H\psi_2 = E_2\psi_2 \quad (3.54)$$

and

$$i\hbar \partial_t \psi_1 = E_1 \psi_1, \quad i\hbar \partial_t \psi_2 = E_2 \psi_2. \quad (3.55)$$

Putting this into our equation above for the superposition (Eq. 3.53) yields

$$\begin{aligned} H(a\psi_1 + b\psi_2 + \dots) &= i\hbar\partial_t(a\psi_1 + b\psi_2 + \dots) && \text{this is Eq. 3.53} \\ \rightarrow aH\psi_1 + bH\psi_2 + \dots &= ai\hbar\partial_t\psi_1 + bi\hbar\partial_t\psi_2 + \dots && \curvearrowright \\ aE_1\psi_1 + bE_2\psi_2 + \dots &= aE_1\psi_1 + bE_2\psi_2 + \dots && \checkmark \end{aligned}$$

Thus, the linear combination  $\psi_{\text{sup}} = a\psi_1 + b\psi_2 + \dots$  is really a solution, too.<sup>44</sup>

This observation is important since we can construct **wave packets** through suitable linear combinations of plane waves.

One possibility is a Gaussian wave packet, where  $A(\vec{p})$  is a Gauss distribution:<sup>45</sup>

$$\begin{aligned} \Psi_{GWP}(\vec{x}, t) &= \int dp^3 A(\vec{p}) e^{i(\vec{p}\vec{x} - Et)/\hbar} \\ &= \int dp^3 \psi_0 e^{i(\vec{p}-\vec{p})^2/4\sigma^2} e^{i(\vec{p}\vec{x} - Et)/\hbar}. \end{aligned}$$

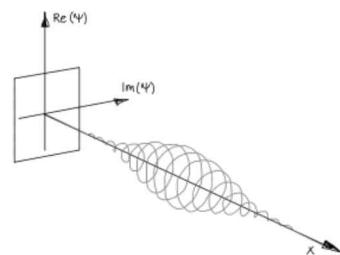
You can see an example of such a Gaussian wave-packet in the margin.

A wave packet is what we use to describe a particle which is localized within some region. By "localized within some region", we mean that we have some idea of the particle's location. In contrast, if we know nothing about a particle's location, it could be anywhere and hence the corresponding wave function is not localized but spreads out all over space. Such a situation can be described by a plane wave which spreads out all over space with equal amplitude. But for real particles that we want to describe in experiments, we always have at least some idea where they are located. Hence, only suitable linear combinations of plane waves (wave packets) are useful to describe real particles.

Take note that solutions of the Schrödinger equation are *complex* functions. Therefore, as promised above, we can understand that physical kets are generally complex.<sup>46</sup> This explains why we need to take the complex conjugate of physical kets to construct the corresponding bras. In addition, it is important to

<sup>44</sup> Again: This only works because there is no term  $\psi^2$  or  $\psi^3$ . In other words, because the Schrödinger equation is linear in  $\psi$ . If you don't believe this, try to do the same calculation with an additional term  $\psi^2$  or  $\psi^3$  in the Schrödinger equation.

<sup>45</sup> Again, we imagine that an integral is something like a finely grained sum.



<sup>46</sup> A physical ket is one that is a solution to the Schrödinger equation.

remember that solutions of the Schrödinger equation only represent probability amplitudes. To get probabilities, we have to take the absolute square of them.

<sup>47</sup> Reminder:  $\lambda = \frac{h}{p}$  and  $E = h\nu$ .

<sup>48</sup> Recall that  $\hbar$  is the reduced Planck constant defined by  $\hbar = h/2\pi$ . We can understand here why we introduced this new constant with an additional factor  $2\pi$ .

Another thing we can now understand is the relations in Eq. 2.2 and Eq. 2.3.<sup>47</sup> In general, a plane wave has the form:  $Ae^{i\frac{2\pi x}{\lambda} - \omega t}$ , where  $\lambda$  is the wavelength and  $\omega$  is the angular frequency of the wave. Comparing this with our solution to the Schrödinger equation in Eq. 3.50 ( $\Psi = e^{i(px-Et)/\hbar}$ ) tells us that<sup>48</sup>

$$\begin{aligned} \frac{2\pi}{\lambda} &\stackrel{\triangle}{=} \frac{p}{\hbar} \\ p &= \frac{\hbar 2\pi}{\lambda} \\ &= \frac{h}{\lambda}. \end{aligned}$$

$\curvearrowright \quad \hbar = \frac{h}{2\pi}$

We also find that

$$\begin{aligned} \omega &\stackrel{\triangle}{=} \frac{E}{\hbar} \\ E &= \hbar\omega \\ &= h\nu, \end{aligned}$$

$\curvearrowright \quad \omega = 2\pi\nu$

where  $\nu$  is the "ordinary" frequency of the wave, while  $\omega$  is the angular frequency.

Before we move on, we should talk about one concept (which we have already used several times) in more detail.

### 3.8 Intermezzo: Eigenvectors and Eigenvalues

The concepts of **eigenvalues** and **eigenvectors** are used throughout the world of quantum mechanics. The eigenvectors,  $\vec{v}$ , and eigenvalues,  $\lambda$ , of a matrix,  $M$ , are exactly those vectors and numbers that fulfill the equation

$$M\vec{v} = \lambda\vec{v}. \quad (3.56)$$

The important thing is that we have *exactly* the same vector  $\vec{v}$  on both sides. This means that the vector  $\vec{v}$  remains, up to a constant  $\lambda$ , unchanged if we multiply it by the matrix  $M$ . For each eigenvector, we have a corresponding eigenvalue.

This is just a definition. Why should we care about eigenvalues and eigenvectors?

Two reasons:

- ▷ We can describe rotations with rotation matrices and the eigenvector of a given rotation matrix is exactly the rotational axis.
- ▷ The structure

$$\text{operator} \times \text{vector} = \text{number} \times \text{same vector}$$

is *exactly* the structure that we use in quantum mechanics.

A matrix is an operator just like the quantum operators we talked about above.<sup>49</sup> The bottom line is that eigenvectors are the states with definite values for the physical quantity represented by the operator. For example, if  $|\Psi_1\rangle$  is an eigenvector of the momentum operator  $\hat{p}$ , we have:  $\hat{p}|\Psi_1\rangle = p_1|\Psi_1\rangle$ .<sup>50</sup>

The corresponding eigenvalue  $p_1$  is the value that we measure. Many operators in quantum mechanics are not represented by matrices, but by differential operators.<sup>51</sup> However, the structure  $\hat{O}|\Psi\rangle = o|\Psi\rangle$  is completely analogous. If  $\hat{O}$  is a differential operator, the only thing that changes is that the eigenvectors become **eigenfunctions**.

There are sophisticated algorithms to calculate the eigenvectors and eigenvalues of a matrix or the eigenfunctions of a differential operator. But a full discussion would lead us too far astray. You can find a full description of these algorithms (at least for matrices) in any linear algebra textbook.<sup>52</sup>

Now it's finally time to return to the angular momentum operator, as promised at the end of Section 3.4.

<sup>49</sup> We will talk about a quantum operator that is indeed represented by a matrix in the following section.

<sup>50</sup> Note that a general ket is not an eigenvector of the momentum operator. For example:

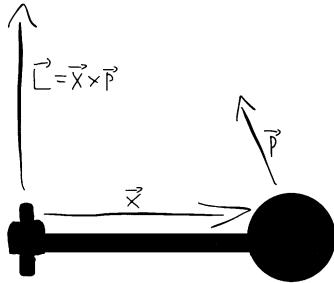
$$\begin{aligned} |\Psi\rangle &= a|\Psi_1\rangle + b|\Psi_2\rangle \\ \Rightarrow \hat{p}|\Psi\rangle &= a\hat{p}|\Psi_1\rangle + b\hat{p}|\Psi_2\rangle \\ &= ap_1|\Psi_1\rangle + bp_2|\Psi_2\rangle \\ &\neq c(a|\Psi_1\rangle + b|\Psi_2\rangle). \end{aligned}$$

The crucial observation is that we don't have the same vector on the left-hand and right-hand side. We can't write  $ap_1|\Psi_1\rangle + bp_2|\Psi_2\rangle$  as some number times the original vector ( $a|\Psi_1\rangle + b|\Psi_2\rangle$ ) since, in general,  $p_1 \neq p_2$ .

<sup>51</sup> Think: something with a derivative such as  $\partial_x$ . Examples are the momentum operator  $-i\hbar\partial_x$  or the energy operator  $i\hbar\partial_t$ .

<sup>52</sup> The part of mathematics that deals specifically with these kinds of problems is known as **spectral theory**.

### 3.9 Angular Momentum



<sup>53</sup> Recall:  $[\hat{p}_i, \hat{x}_j] = -i\hbar\delta_{ij}$

In Section 3.4.2, I mentioned that the angular momentum operator follows directly from the classical angular momentum  $\vec{L} = \vec{x} \times \vec{p}$  if we replace  $\vec{p}$  with the corresponding momentum operator  $-i\hbar\vec{\partial}$ .

In addition, we learned in Section 3.5 that the order in which we measure position and momentum affects the outcome. This curious property of quantum mechanics is encoded in the canonical commutation relation (Eq. 3.42)<sup>53</sup>.

We can find a very similar relation for the angular momentum operator. Using the explicit form of the angular momentum operator,<sup>54</sup>

$$\begin{aligned}\hat{L}_i &= \epsilon_{ijk}\hat{x}_j\hat{p}_k \\ &= \epsilon_{ijk}\hat{x}_j(-i\hbar\partial_k) \\ &= -i\hbar\epsilon_{ijk}\hat{x}_j\partial_k,\end{aligned}\quad \begin{array}{l} \curvearrowright \\ \curvearrowright \end{array} \quad \text{Eq. 3.38}$$

we can calculate the **angular momentum commutation relation**:<sup>55</sup>

$$[\hat{L}_i, \hat{L}_j] = i\hbar\epsilon_{ijk}\hat{L}_k \quad (3.57)$$

<sup>54</sup> It is convenient to write the operator in index notation. In index notation, the cross product  $A \times B$  reads as  $\epsilon_{ijk}A_iB_j$ , where  $\epsilon_{ijk}$  is the Levi-Civita symbol. For the definition of the Levi-Civita symbol, see Eq. 2.

<sup>55</sup> Explicitly this equation reads, for example,  $[\hat{L}_x, \hat{L}_y] = i\hbar\hat{L}_z$  and  $[\hat{L}_y, \hat{L}_z] = i\hbar\hat{L}_x$  and  $[\hat{L}_z, \hat{L}_x] = i\hbar\hat{L}_y$ . To calculate this you have to remember that there is an implicit ket  $|\Psi\rangle$  behind each term which we are too lazy to write all the time.

This equation is also known as the **angular momentum algebra**. The crucial point is that measurements of the angular momentum along *different* axes affect each other. In particular, a measurement of the angular momentum along the  $x$ -axis affects the angular momentum along the  $y$ - and  $z$ -axes. This is in contrast to the momentum and location operators, wherein measurements along *different* axes do not affect each other since, for example,  $\partial_x y = 0$  and  $\partial_x \partial_y = \partial_y \partial_x$ .

In practice, this means that we can't determine the angular momentum of a given particle along different axes with arbitrary precision. Each time we measure it along one axis, we lose information about the angular momentum along other axes.

However, the craziness of angular momentum in quantum mechanics does not stop here. Next, we'll talk about an important subtlety of the angular momentum operator in quantum mechanics. As we will see in a moment, this subtlety leads us to a new and incredibly important quantity. Unfortunately, the full story is far too long to include it here. If you are interested, you can find all of the details in my book "Physics from Symmetry".<sup>56</sup> The following section will only sketch the full story.

<sup>56</sup> Jakob Schwichtenberg. *Physics from Symmetry*. Springer, Cham, Switzerland, 2018. ISBN 978-3319666303

## 3.10 Spin

For many systems, we can use the "naive" operator obtained by using  $\vec{L} = \vec{x} \times \vec{p}$  and replacing  $\vec{p}$  with  $-i\hbar\vec{\partial}$ . The resulting operator is exactly the generator of rotations. For example, let's say we have a function  $f(\vec{x})$  and want to describe it in a rotated coordinate system. We can achieve an infinitesimal rotation as follows:

$$f(\vec{x}) \rightarrow (1 + \epsilon \hat{L}) f(\vec{x}).$$

However, this is not always the full story.

Sometimes, we need to describe our quantum systems with objects that have more than one component:<sup>57</sup>

$$\Psi(x, t) = \begin{pmatrix} \Psi_1(x, t) \\ \Psi_2(x, t) \end{pmatrix}.$$

A crucial point is that each component of these objects is a *function* of  $x$  and  $t$ . This means that acting on such objects with symmetry generators results in two things:<sup>58</sup>

1. On one hand, they change the spatial and temporal coordinates  $x \rightarrow 1 + Gx$ , where  $G$  denotes a generator. This is exactly the kind of transformation we talked about previously:  $\Psi(x) \rightarrow \Psi(x + \epsilon)$  (Eq. 3.37).
2. But on the other hand, transforming multi-component objects can also mix these components:

<sup>57</sup> Think: like a vector. The reason why we need such objects with more than one component is that they are necessary to describe certain quantum systems. For example, you maybe already know that electromagnetic waves can be polarized. This means that there are internal degrees of freedom and we therefore need objects with more than one component to describe them. As already mentioned above, electromagnetic waves are transmitted by particles (photons) and therefore we need objects with several components to describe these quantum particles. We will talk about another example in a moment.

<sup>58</sup> For the other operators there is no such subtlety since translations do not mix components.

$$\begin{pmatrix} \Psi_1(x, t) \\ \Psi_2(x, t) \end{pmatrix} \rightarrow \begin{pmatrix} \Psi_2(x, t) \\ \Psi_1(x, t) \end{pmatrix}$$

<sup>59</sup> An example: We want to look at a vector with four components

$$A_\mu = \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix} \text{ from a different}$$

perspective. Formulated differently, we want to describe it in a rotated coordinate system. The result could

$$\text{be something like } A'_\mu = \begin{pmatrix} A'_0 \\ A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \begin{pmatrix} A_0 \\ -A_2 \\ A_1 \\ A_3 \end{pmatrix}. A'_\mu \text{ and } A_\mu \text{ describe the}$$

same field in coordinate systems that are rotated by 90° around the z-axis relative to each other. Through the transformation the components get mixed.

<sup>60</sup> Think: Something with a derivative  $\partial_i$  in it.

<sup>61</sup> Orbital angular momentum is an important quantity, for example, when an object revolves around a second one. Think: Earth and the sun.

<sup>62</sup> Think along the lines of a spinning ball, although you shouldn't take this picture too seriously. This is in contrast to orbital angular momentum which only exists when two objects revolve around each other.

<sup>63</sup> The correct name for the two-component objects that we deal with in quantum mechanics is **spinors**. Spinors are somewhat strange objects and not simply two-dimensional vectors. However, a full discussion and a derivation of these generators is too long to include it here. If you're interested you can find all the details in my book "Physics from Symmetry".

The most familiar example is when we rotate a vector. Under a rotation, the vector components get mixed.<sup>59</sup> And when the components depend on the position  $x$ , these functions are modified, too.

The generator that causes the mixing of the components is a *matrix*. The generator that causes the rotation of the argument of a function is a *differential operator*.<sup>60</sup>

Our goal is to find a general quantum operator that describes angular momentum. Identifying the quantum operators for energy and linear momentum (and the corresponding generators) was relatively straightforward.

However, we now have a generator composed of two parts. Which part is the correct one?

Actually, both! For the angular momentum that we know from classical mechanics, the naive operator  $\vec{L} = \vec{x} \times (-i\hbar\vec{\partial})$  is correct. To avoid confusion, this type of angular momentum is referred to as **orbital angular momentum**.<sup>61</sup> The second, equally important operator is identified with the generator responsible for mixing the vector's components (i.e., a matrix). Since both quantities are required when we consider rotations, they can't be too different. In fact, the quantity described by this second operator is commonly interpreted as *internal angular momentum*, or spin.<sup>62</sup> The total angular momentum of the system is simply the sum of the orbital angular momentum and the spin. Specifically, the correct generators of spin for two-component states<sup>63</sup> are

$$\hat{S}_i = \frac{\hbar}{2} \sigma_i, \quad (3.58)$$

where  $\sigma_i$  are  $(2 \times 2)$  matrices known as **Pauli matrices**:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.59)$$

Here we use the labels 1, 2, and 3 instead of  $x$ ,  $y$ , and  $z$ . For example, the operator  $\hat{S}_3$  reads

$$\hat{S}_3 = \frac{\hbar}{2} \sigma_3 = \begin{pmatrix} \frac{\hbar}{2} & 0 \\ 0 & -\frac{\hbar}{2} \end{pmatrix}. \quad (3.60)$$

The corresponding eigenvectors are

$$v_{\frac{\hbar}{2}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v_{-\frac{\hbar}{2}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (3.61)$$

with eigenvalues  $\frac{\hbar}{2}$  and  $-\frac{\hbar}{2}$ , respectively. These eigenvectors are our basic spin states. The eigenvalues are the values that we measure. Let's say our system is in a state described by

$$\Psi(x, t) = \begin{pmatrix} \psi(x, t) \\ 0 \end{pmatrix}, \quad (3.62)$$

and we want to measure the spin in the 3-direction. Therefore, we calculate

$$\begin{aligned} \hat{S}_3 \Psi(x, t) &\stackrel{\text{Eq. 3.62}}{=} \hat{S}_3 \begin{pmatrix} \psi(x, t) \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{\hbar}{2} & 0 \\ 0 & -\frac{\hbar}{2} \end{pmatrix} \begin{pmatrix} \psi(x, t) \\ 0 \end{pmatrix} \\ &= \frac{\hbar}{2} \begin{pmatrix} \psi(x, t) \\ 0 \end{pmatrix}. \end{aligned} \quad \begin{matrix} \curvearrowleft & \text{Eq. 3.60} \\ & \curvearrowleft \end{matrix}$$

This is exactly the structure we talked about in the previous sections. We act on the object that describes our state with the operator and get back what we will measure in an experiment. In this case, we would measure the value  $\hbar/2$  for the spin in the 3-direction.

Systems are usually in a superposition of different spin states. A convenient notation is  $|\uparrow\rangle$  for the state with spin  $\hbar/2$  and  $|\downarrow\rangle$  for the state with spin  $-\hbar/2$ .<sup>64</sup> A general state can therefore be expressed as

$$|\Psi\rangle = a |\uparrow\rangle + b |\downarrow\rangle.$$

As before, the coefficients  $a$  and  $b$  are directly related to the probability of obtaining spin values of  $\frac{\hbar}{2}$  and  $-\frac{\hbar}{2}$ , respectively<sup>65</sup>

<sup>64</sup> The state with spin  $\hbar/2$  is usually called "spin up" state and the state with spin  $-\hbar/2$  "spin down" state.

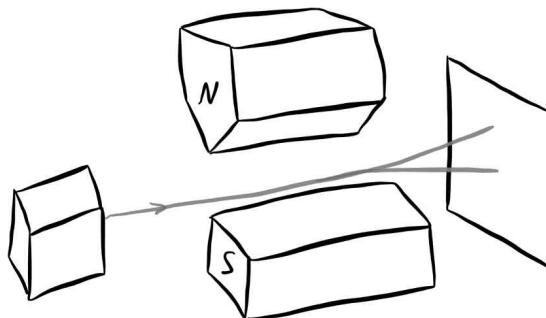
<sup>65</sup> Explicitly,  $|a|^2$  is the probability of measuring  $\hbar/2$  and  $|b|^2$  the probability of measuring  $-\hbar/2$ .

Now comes the crucial point.

In contrast to the classical mechanics regime, under which angular momentum can assume any value, quantum mechanical spin can only take on one of two values. Either we measure the value  $\hbar/2$ , or we measure the value  $-\hbar/2$ . There is nothing in between. In other words:

**Spin is quantized!**

Historically this was a huge surprise for every physicist. The quantization of angular momentum was discovered experimentally and not predicted by any theorist. The discovery was made by the second most famous quantum experiment:<sup>66</sup> the **Stern-Gerlach experiment**. The experimental setup is shown below.<sup>67</sup>



Classically, we would expect a continuous distribution of atoms on the screen since particles with different spin alignment, relative to the magnetic field, get deflected differently. But what was observed is that the atoms ended up in just two distinct regions. This indicates that only two spin orientations relative to the magnetic fields are possible.

Today, with the power of hindsight (and a more complete understanding of quantum mechanics), this result is no longer completely mysterious. As we have seen above, this result can be derived using techniques similar to those we used to derive other quantum features.

<sup>66</sup> The most famous quantum experiment is the double slit experiment that we already talked about in Section 2.1.

<sup>67</sup> Historically silver atoms were used for technical reasons. The angular momentum of silver atoms depends dominantly on the angular momentum of a single electron. Since an experiment with single electrons was too difficult at the time, this was the next best thing they could do. Nowadays, physicists are able to repeat the experiment with single electrons. The result is the same.

However, this is not the only curious property of spin. We already learned that measurements of momentum and location along the *same* axis affect each other. We also learned that measurements of angular momentum along *different* axes affect each other. Now we should check to see how repeated spin measurements along various axes affect the outcomes we obtain.

Mathematically, these properties are encoded in the canonical commutation relation (Eq. 3.42)<sup>68</sup> and the angular momentum commutation relation (Eq. 3.57)<sup>69</sup>.

Using the explicit matrices for the spin operators (Eq. 3.58), we can calculate the commutation relations as:

$$[\hat{S}_i, \hat{S}_j] = i\hbar\epsilon_{ijk}\hat{S}_k . \quad (3.63)$$

For example, we have

$$\begin{aligned} [\hat{S}_1, \hat{S}_2] &= \hat{S}_1\hat{S}_2 - \hat{S}_2\hat{S}_1 \\ &= \begin{pmatrix} 0 & \frac{\hbar}{2} \\ \frac{\hbar}{2} & 0 \end{pmatrix} \begin{pmatrix} 0 & -i\frac{\hbar}{2} \\ i\frac{\hbar}{2} & 0 \end{pmatrix} \quad \text{Eq. 3.58, Eq. 3.59} \\ &\quad - \begin{pmatrix} 0 & -i\frac{\hbar}{2} \\ i\frac{\hbar}{2} & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{\hbar}{2} \\ \frac{\hbar}{2} & 0 \end{pmatrix} \\ &= \begin{pmatrix} i\frac{\hbar^2}{4} & 0 \\ 0 & -i\frac{\hbar^2}{4} \end{pmatrix} \quad \text{matrix products} \\ &\quad - \begin{pmatrix} -\frac{i\hbar^2}{4} & 0 \\ 0 & \frac{i\hbar^2}{4} \end{pmatrix} \\ &= \begin{pmatrix} \frac{i\hbar^2}{2} & 0 \\ 0 & -\frac{i\hbar^2}{2} \end{pmatrix} \\ &= i\hbar\hat{S}_3 \quad \text{Eq. 3.58, Eq. 3.59} \\ &= i\hbar\epsilon_{12k}\hat{S}_k \quad \epsilon_{12k} = 0 \text{ except for } k = 3 \\ &= i\hbar\epsilon_{123}\hat{S}_3 . \end{aligned}$$

<sup>68</sup> Reminder:  $[\hat{p}_i, \hat{x}_j] = -i\hbar\delta_{ij}$  and the square brackets denote, in general, commutators:  $[A, B] = AB - BA$ . A commutator encodes if it makes a difference if we first apply  $B$  and then  $A$  or first  $A$  and then  $B$ . Formulated differently, a non-zero commutator indicates that it makes a difference in what order we apply two operators.

<sup>69</sup> Reminder:  $[\hat{L}_i, \hat{L}_j] = i\hbar\epsilon_{ijk}\hat{L}_k$ .

$\epsilon_{ijk}$  denotes the totally antisymmetric Levi-Civita symbol.

The general commutation relation in Eq. 3.63 is known as the **spin algebra**. It tells us, for example, that it makes a difference whether we first measure the spin in the  $x$  direction and then in the  $y$  direction or the other way round. Take note that this relation is completely analogous to the angular momentum algebra, which provides additional evidence that spin is some type of angular momentum. We will talk a bit more about spin in Section 10.1.

An important side-note is that not every object has spin. However, almost all elementary particles have a non-zero spin. The only exception is the recently observed Higgs boson. For electrons (and also, for example, protons, neutrons or quarks), we need two-component objects. We say they have spin  $1/2$ , for reasons that we will discuss in more detail in Section 10.1. Other elementary particles (photons, for example) have spin  $1$ , and we need four-component objects (four-vectors) to describe them. The corresponding spin operators that allow us to extract the spin values of photons are therefore  $(4 \times 4)$  matrices.

---

So what we have learned in the previous sections is that in quantum mechanics, it's impossible to know the location and the momentum at the same time with arbitrary precision. Moreover, we have learned that we can't know the full angular momentum vector  $\vec{L} = (L_x, L_y, L_z)^T$ , since each time we measure one component, we lose information about the other components.<sup>70</sup>

<sup>70</sup> This is what the angular momentum commutation relation (Eq. 3.57) and the canonical commutation relation (Eq. 3.42) tell us.

Thus, we have to be a bit more careful in quantum mechanics which quantities we use to describe our systems and this is what the next section is about.

### 3.11 Quantum Numbers

In classical mechanics, we simply use a vector  $\vec{L} = (L_x, L_y, L_z)^T$  to describe the angular momentum of an object. Similarly, we usually specify the location *and* the momentum of objects to describe them. In quantum mechanics, we can't know  $L_x$ ,  $L_y$  and  $L_z$  at the same time with arbitrary precision. At most, we can know *one* of them precisely. We pick just one component of the angular momentum and use it to specify the state of our system. Conventionally, we choose  $L_z$ . There is no physical reason behind this since it is completely our choice what we call the  $z$ -axis and what the  $x$ -axis or  $y$ -axis.

However, it turns out that there is a bit more we can know about the angular momentum of a given system. We can also measure the total angular momentum in addition to one component! Classically, the length of the angular momentum vector (squared) is given by  $\vec{L}^2 = L_x^2 + L_y^2 + L_z^2$ . This quantity tells us the total magnitude of the angular momentum vector. The corresponding quantum operator is defined analogously

$$\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2. \quad (3.64)$$

This quantity is important because we can measure it without affecting the individual components. Again, this property is encoded in a commutation relation:  $[\hat{L}^2, \hat{L}_i] = 0$ . In words, this means that we can know  $\hat{L}^2$  and one  $\hat{L}_i$  at the same time with arbitrary precision<sup>71</sup>.

<sup>71</sup> Mathematically, we say that  $\hat{L}^2$  commutes with  $\hat{L}_x$ ,  $\hat{L}_y$  and  $\hat{L}_z$ .

So to summarize, we specify the angular momentum of a given quantum system using the  $z$ -component of the angular momentum,  $\hat{L}_z$ , and the total angular momentum,  $\hat{L}^2$ . It is conventional to use the label  $m$  for the angular momentum in the  $z$ -direction:

$$\hat{L}_z |m\rangle = \hbar m |m\rangle$$

and the label  $l$  for the total angular momentum:<sup>72</sup>

$$\hat{L}^2 |l\rangle = \hbar^2 l(l+1) |l\rangle .$$

<sup>72</sup> Don't let yourself get confused why we have  $l(l+1)$  instead of simply  $l$ . There is a long story behind this, but it's mainly just a convention.

<sup>73</sup> For example, for the hydrogen atom we have  $E_n = -\frac{1}{2}m_e \left(\frac{e^2}{4\pi\epsilon_0\hbar}\right)^2 \frac{1}{n^2}$  or for a particle in a box we have  $E_n = \frac{n^2\pi^2\hbar^2}{D^22m}$ .

In addition, the energy operator  $\hat{H}$  often commutes with  $\hat{L}_z$  and  $\hat{L}^2$ , so we can use it as a third label. The conventional label for the energy eigenvalues is  $n$ .<sup>73</sup>

So in summary: We often label our states using the three labels  $m, l$  and  $n$ :  $|n, m, l\rangle$ . Labels like this are known as **quantum numbers**. It is somewhat an art to pick the right quantum numbers for a given problem. However, the basic task is always to find operators that commute with each other,  $[\hat{A}, \hat{B}] = 0$ , since precise and simultaneous measurements are only possible if the operators commute.

The whole business becomes even more complicated if we are dealing with a particle with spin. Then, we have in addition to the labels above an additional label for the spin in the  $z$ -direction and one for the total spin.<sup>74</sup> In addition, for some systems, it is important to determine the relative contributions of the spin and orbital angular momentum to the total angular momentum.

Similarly, we have to decide between position and momentum labels since  $[\hat{p}, \hat{x}] \neq 0$ . Our choice always depends on the problem and experiment at hand. Sometimes we pick the momentum and sometimes the location. However, we can never use both labels at the same time.<sup>75</sup>

<sup>74</sup> We have these additional labels since the spin operators commute with the angular momentum operators  $[\hat{S}_i, \hat{L}_j] = 0$ . In other words, we can measure the orbital angular momentum and spin at the same time with arbitrary precision.

<sup>75</sup> Take note, for example, that the wave functions we discussed in Section 3.3 are only functions of  $x$  and  $t$ . However, we can switch the labels through a Fourier transform. The result of such a transformation is a  $\Psi(p, t)$ , i.e., a wave function that only depends on the momentum. For more on this, see Appendix B.

Next, we turn to an important question that remains unanswered: how is quantum mechanics connected to classical mechanics?

Both yield excellent results which agree with what we observe in nature, however for different kinds of systems. Quantum mechanics describes the behavior of elementary particles, whereas classical mechanics describes how macroscopic objects behave. But a macroscopic object like a ball also consists of (many, many) elementary particles. So it therefore must be pos-

sible to derive the laws of classical mechanics using the laws of quantum mechanics by averaging over many many elementary particle systems. This is what the next section is about.

