Chapter 2
Polarization

In this chapter, we make the transition from classical mechanics to quantum mechanics by considering light polarization. This leads us directly to two key concepts of quantum mechanics, vector space and probability. For the first time, we encounter the problem of measurement in quantum mechanics.

The approach to quantum mechanics in the preceding chapter is based on the description of the time evolution of a state by means of a differential equation. In this chapter, we choose a different approach. We consider (for now) not the Schrödinger equation or another description of the space-time behavior, but instead the emphasis is now on how we can define *states* (for the moment, time-independent states).

Again we start from classical formulations which we 'pep up' quantum mechanically. For this purpose we first show that under certain circumstances, we can treat electromagnetic waves as if they propagate in a two-dimensional *complex vector space*.[1] As is known from optics, we can express the intensities of light waves as the absolute squares of their amplitudes. After reviewing classical formulations, we extend these ideas to the quantum-mechanical case by means of a reinterpretation which is, while not mandatory, very plausible. In this interpretation, the amplitudes do not lead to intensities, but instead to *probabilities*. At this point we will see for the first time that the concept of *measurement* in quantum mechanics is not as trivial as it is in classical mechanics.

[1] If this term is not familiar (or forgotten): The basic concepts are summarized in Appendix G, Vol. 1. In addition, we will return to this topic in Chap. 4. For the moment, it is enough to know that e.g. the set of all vectors $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ with $a_i \in \mathbb{C}$ forms a two-dimensional complex vector space. An important property is that any linear combination of two vectors is itself a valid vector in this space.

J. Pade, *Quantum Mechanics for Pedestrians 1*, Undergraduate Lecture Notes in Physics, https://doi.org/10.1007/978-3-030-00464-4_2

We will base our discussion on the *polarization* of light, which should be familiar from lectures and lab courses.[2]

2.1 Light as Waves

We first derive the 'minimal description' of a classical electromagnetic wave. Using the common definition of linear and circular polarization, we see that we can describe these 'typical waves' in a two-dimensional complex vector space.

2.1.1 The Typical Shape of an Electromagnetic Wave

We start with the description of an electromagnetic plane wave as[3]

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{i(\mathbf{kr}-\omega t)};\ \mathbf{B} = \frac{\mathbf{k} \times \mathbf{E}}{c} \tag{2.1}$$

with $\mathbf{k} \cdot \mathbf{E}_0 = 0$ (a transverse wave, as follows from the first Maxwell equation. In a charge-free region of space, it states that $\nabla \mathbf{E} = 0$); $\omega^2 = c^2 \mathbf{k}^2$ (dispersion relation for zero rest mass), with $\mathbf{E}_0 \in \mathbb{C}^3$. In the following, we restrict our considerations to the electric field[4] $\mathbf{E}$; the magnetic field can be calculated from $\mathbf{E}$ by using (2.1).

It holds quite generally that the description of a plane wave can be made considerably simpler and more transparent by means of a suitable choice of the coordinate system, without losing any of its physical significance. We choose the new z axis to point in the direction of propagation of the wave, i.e. the $\mathbf{k}$ direction—in other words, $\mathbf{k} = (0, 0, k)$—and obtain

$$\mathbf{E}(\mathbf{r}, t) = \left(E_{0x}, E_{0y}, 0\right) e^{i(kz-\omega t)}. \tag{2.2}$$

The z component disappears due to the transverse nature of the wave (see exercises at the end of this chapter).

---

[2] From the theory of electromagnetism, we know that light is a transverse wave, i.e. that its electric field oscillates perpendicular to its direction of propagation. The *polarization* describes the orientation of this oscillation.

Polarization is often regarded as an esoteric and specialized topic, possibly because we cannot see directly whether light is polarized. However, it is a ubiquitous phenomenon in our environment—natural light is almost always polarized, at least partially. Many animals, such as bees or other insects, take advantage of this; they can detect and analyze light polarization. In our daily life, polarization is used e.g. in polarizing filters for cameras or some sunglasses. Moreover, the fundamentals of the formal treatment of polarization are also very simple, as we shall see below.

[3] We note that a real light wave is only approximately described by a plane wave, since that would have the intensity at all points and all times. However, this approximate description is common for several reasons, and suffices for our purposes here.

[4] In this connection also called the *light vector*.

The amplitude can be written quite generally as

$$E_{0x} = e^{i\alpha} |E_{0x}| ; \; E_{0y} = e^{i\beta} \left| E_{0y} \right| ; \; \alpha, \beta \in \mathbb{R}. \tag{2.3}$$

It follows that

$$\mathbf{E}(\mathbf{r}, t) = \left(|E_{0x}|, \left| E_{0y} \right| e^{i(\beta-\alpha)}, 0 \right) e^{i(kz-\omega t+\alpha)}. \tag{2.4}$$

We can now put $\alpha = 0$ without loss of generality, since the last equation shows that any value of α can be compensated by a suitable choice of the zero of time. In order to avoid confusion, we rename β as δ. Then the typical form of an electromagnetic wave is given by[5]:

$$\mathbf{E}(\mathbf{r}, t) = \left(|E_{0x}|, \left| E_{0y} \right| e^{i\delta}, 0 \right) e^{i(kz-\omega t)}. \tag{2.5}$$

2.1.2 Linear and Circular Polarization

For purposes of illustration (we will include some figures in the following), we consider in this subsection only the real part of the wave function (the imaginary part alone would be just as suitable):

$$E_x(\mathbf{r}, t) = |E_{0x}| \cos(kz - \omega t) ; \; E_y(\mathbf{r}, t) = \left| E_{0y} \right| \cos(kz - \omega t + \delta). \tag{2.6}$$

$\delta \in \mathbb{R}$ can assume all possible values. One can, however, single out two basic cases, namely $\delta = 0$ (*linear polarization*) and $\delta = \pm\pi/2$ (*elliptical or circular polarization*).

2.1.2.1 Linear Polarization

With the choice $\delta = 0$, we have

$$E_x(\mathbf{r}, t) = |E_{0x}| \cos(kz - \omega t) ; \; E_y(\mathbf{r}, t) = \left| E_{0y} \right| \cos(kz - \omega t). \tag{2.7}$$

It follows immediately that

$$E_y = \frac{\left| E_{0y} \right|}{|E_{0x}|} E_x. \tag{2.8}$$

This is a straight line on which the light vector oscillates back and forth—hence the name *linear* polarization, see Fig. 2.1.

Basic types of this polarization are obtained by setting one component equal to zero:

[5] The relative phase could of course be associated with the y component instead of the x component.

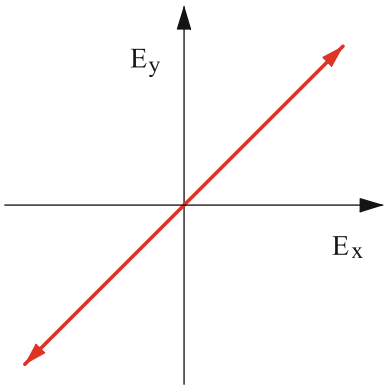

Fig. 2.1 Linear polarization. The z-axis points out of the image plane

$$\begin{aligned} &\text{horizontally polarized:} \quad E_x(\mathbf{r},t) = |E_{0x}|\cos(kz-\omega t)\,; \;\; E_y = \left|E_{0y}\right| = 0 \\ &\text{vertically polarized:} \quad E_x = |E_{0x}| = 0; \;\; E_y(\mathbf{r},t) = \left|E_{0y}\right|\cos(kz-\omega t)\,. \end{aligned} \tag{2.9}$$

The names are self-explanatory. Due to the vector character of the electric field, it follows readily that any linearly-polarized wave can be written as a superposition of horizontally- and vertically-polarized waves.

2.1.2.2 Elliptical and Circular Polarization

In this case, we choose $\delta = \pm\pi/2$ and this means that

$$\begin{aligned} E_x(\mathbf{r},t) &= |E_{0x}|\cos(kz-\omega t) \\ E_y(\mathbf{r},t) &= \left|E_{0y}\right|\cos(kz-\omega t \pm \pi/2) = \mp\left|E_{0y}\right|\sin(kz-\omega t)\,. \end{aligned} \tag{2.10}$$

It follows from this that

$$\left(\frac{E_x}{|E_{0x}|}\right)^2 + \left(\frac{E_y}{\left|E_{0y}\right|}\right)^2 = 1. \tag{2.11}$$

The arrowhead of the light vector thus moves on an ellipse with semiaxes $|E_{0x}|$ and $\left|E_{0y}\right|$—hence the name *elliptical* polarization; see Fig. 2.2. The direction of rotation can be determined by using

$$\tan\vartheta = \frac{E_x}{E_y} = \mp\frac{\left|E_{0y}\right|}{|E_{0x}|}\tan(kz-\omega t) = \frac{\left|E_{0y}\right|}{|E_{0x}|}\tan(\pm\omega t \mp kz) \tag{2.12}$$

(most easily seen for fixed z).

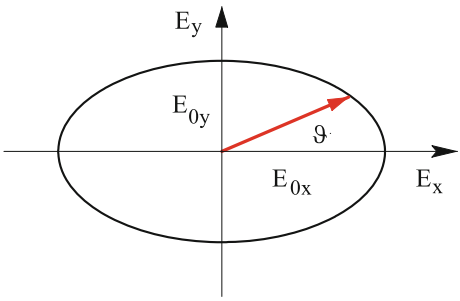

Fig. 2.2 Elliptical polarization. The z-axis is directed out of the image plane

In particular, for $|E_{0x}| = |E_{0y}|$, the ellipse becomes a circle and we have circularly-polarized light, i.e. right circularly-polarized light with the upper sign, left circularly-polarized light with the lower sign.[6]

2.1.3 From Polarization to the Space of States

In summary, we have in the complex representation (remember $e^{\pm i\pi/2} = \pm i$) for the linearly (horizontal h/vertical v) and circularly (right r/left l) polarized waves:

$$\begin{aligned}
\mathbf{E}_h &= (|A_{0x}|, 0, 0)\, e^{i(kz-\omega t)} \\
\mathbf{E}_v &= (0, |B_{0x}|, 0)\, e^{i(kz-\omega t)} \\
\mathbf{E}_r &= (|C_{0x}|, i\,|C_{0x}|, 0)\, e^{i(kz-\omega t)} \\
\mathbf{E}_l &= (|C_{0x}|, -i\,|C_{0x}|, 0)\, e^{i(kz-\omega t)}.
\end{aligned} \tag{2.13}$$

So far, we have just repeated material that should be known from previous semesters. Now we turn to something that is (quite possibly) new. We begin by noting that the representation (2.13) is redundant and we can simplify it further.

2.1.3.1 Simplifying the Notation

To achieve this simplification, we have to restrict our world: it will consist exclusively of the waves given by (2.13). In particular, there is e.g. no other direction of propagation and no other wave number k. Then we can simplify as follows:

1. The factor $e^{i(kz-\omega t)}$ occurs everywhere, so we can omit it.
2. Since the third component is always zero, we suppress it. In other words, our little world is *two dimensional*.

[6] In physical optics, right and left circular polarization is usually defined the other way around (optics convention).

3. The notation as a row vector was chosen for typographical convenience; the correct notation is as a column vector.[7]
4. We fix the undetermined quantities $|A_{0x}|$ etc. in such a way that the respective vector has length 1, and thus represents a unit vector. We can then build up a general vector by taking appropriate linear combinations of these unit vectors.

In summary:

$$\begin{aligned}(|A_{0x}|,0,0)\,e^{i(kz-\omega t)} &\overset{1.}{\to} (|A_{0x}|,0,0)\\ \overset{2.}{\to} (|A_{0x}|,0) &\overset{3.}{\to} \begin{pmatrix}|A_{0x}|\\0\end{pmatrix} \overset{4.}{\to} \begin{pmatrix}1\\0\end{pmatrix}\end{aligned} \tag{2.14}$$

and[8]

$$\begin{aligned}(|C_{0x}|,i\,|C_{0x}|,0)\,e^{i(kz-\omega t)} &\overset{1.}{\to} (|C_{0x}|,i\,|C_{0x}|,0)\\ \overset{2.}{\to} (|C_{0x}|,i\,|C_{0x}|) &\overset{3.}{\to} \begin{pmatrix}|C_{0x}|\\i\,|C_{0x}|\end{pmatrix} \overset{4.}{\to} \tfrac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix}.\end{aligned} \tag{2.15}$$

In short, we go from a three-dimensional to a two-dimensional *complex vector space* which we call the *state space*. In this space, the states associated with the vectors (2.13) are written as *two-component vectors* which depend neither on position nor on time. For a convenient shorthand notation comparable to $\mathbf{E}_h$ in (2.13), we introduce the notation $|h\rangle$, $|v\rangle$, $|r\rangle$ and $|l\rangle$ for light in the linear horizontal, linear vertical, right circular and left circular polarized states, respectively. For these states we have the *representation* (denoted by the symbol $\cong$)

$$\begin{aligned}|h\rangle \cong \begin{pmatrix}1\\0\end{pmatrix};\ |v\rangle \cong \begin{pmatrix}0\\1\end{pmatrix}\\ |r\rangle \cong \tfrac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix};\ |l\rangle \cong \tfrac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}.\end{aligned} \tag{2.16}$$

We emphasize that in our 'small world' this representation is completely equivalent to the representation in (2.13).

A note concerning the symbol $\cong$: actually, the representation (2.16) is just one of infinitely many possible ones, based on the fact that in (2.13), we identify the horizontal direction with the x axis. Of course, this is not to be taken for granted, since the y axis might just as well play the same role, given a corresponding orientation. Generally, one can start from any representation $|h\rangle \cong \frac{1}{\sqrt{a^2+b^2}}\begin{pmatrix}a\\b\end{pmatrix}$ and

[7] In the following, we want to multiply vectors by *matrices*. In the usual notation, a matrix acts on a vector from the the left, which therefore—according to the usual rules of matrix multiplication—must be a column vector. See also Appendix F, Vol. 1, on linear algebra.

[8] The length of the vector $\begin{pmatrix}1\\i\end{pmatrix}$ is given by $\sqrt{2}$; we explain the reasoning for this in Chap. 4.

$|v\rangle \mathrel{\widehat{=}} \frac{1}{\sqrt{c^2+d^2}} \begin{pmatrix} c \\ d \end{pmatrix}$ with $a^*c + b^*d = 0$. We therefore identify specific representations by the special symbol $\mathrel{\widehat{=}}$ and do not simply use the equals sign.[9]

The question may arise as to whether the representation (2.16) is not oversimplified. In this context, we recall the following: In physics, the objective of the formal description is not to describe 'nature' (whatever is meant by this term) *directly* but rather to find a *model* for a part of nature and to describe this model as accurately as possible. This is also reflected in the much-quoted 'accuracy' of the natural sciences. It is not the description of nature which is exact, but at most the formal treatment of the model (if at all).[10] Kepler's laws, for example, do not describe the conditions in the solar system exactly, as is well known: the planets influence each other, they are not point masses, there are moons and the solar wind, etc. Kepler's laws, however, are exact within the framework of the model 'point mass earth moves around point mass sun', and this model is correct and sufficiently precise for many applications.[11]

In this sense, the description by models is not unique, but depends on the particular question being considered. The general rule is: as easy as possible, as elaborate as necessary.[12] This is easily said, but of course it is not clear from the outset in all cases what it means in detail. In fact, it is *the* art in science to carve out meaningful, workable models from the 'jumble of reality', neither oversimplified nor overcomplicated.

For the following considerations, our quite modest, simplistic representation (2.16) will be sufficient: We need no direction of propagation, no plane waves, no explicit time behavior and so on.

2.1.3.2 Two Basic Systems

With $|h\rangle$, $|v\rangle$ and $|r\rangle$, $|l\rangle$ we have two pairs of linearly-independent vectors and therefore two basis systems for our two-dimensional vector space. They can be transformed into each other by

---

[9] Different symbols are in use to denote representations; Fließbach writes :=, for example. Apart from that, many authors denote representations not by a special symbol, but by simply writing =.

[10] Also, the general mathematical modelling uses concepts that are implemented only approximately in reality. A time-honored example is Euclidean geometry with its points and lines, which strictly speaking do not exist anywhere in our real world. Yet no one doubts that Euclidean geometry is extremely useful for practical calculations. "Although this may be seen as a paradox, all exact science is dominated by the idea of approximation" (Bertrand Russell).

[11] Most theoretical results are based on approximations or numerical calculations and are in this sense not strictly precise. This naturally applies a fortiori to experimental results. Even though there are high-precision measurements with small relative errors of less than a part per billion, it has to be noted that *each* measurement is inaccurate. Nevertheless, one can estimate this inaccuracy in general quite precisely; keyword 'theory of errors'.

[12] If several theories describe the same facts, one should prefer the simplest of them (this is the principle of parsimony in science, also called Occam's razor: "*entia non sunt multiplicanda praeter necessitatem*").

$$\begin{aligned} |r\rangle &= \tfrac{|h\rangle + i|v\rangle}{\sqrt{2}} \\ |l\rangle &= \tfrac{|h\rangle - i|v\rangle}{\sqrt{2}} \end{aligned} \tag{2.17}$$

and

$$\begin{aligned} |h\rangle &= \tfrac{|r\rangle + |l\rangle}{\sqrt{2}} \\ |v\rangle &= \tfrac{|r\rangle - |l\rangle}{i\sqrt{2}} \end{aligned} \tag{2.18}$$

These relations hold *independently* of the representation, and this is why we write $=$ here, and not $\cong$. Mathematically, (2.17) and (2.18) are basis transformations; physically, these equations mean that we can consider linearly-polarized light as a superposition of right and left circularly-polarized light—and of course vice versa.

2.1.3.3 Intensity and the Absolute Square Amplitude

If we send right circularly-polarized light through an analyzer (linearly horizontal/vertical), then the relative intensity of horizontally and vertically polarized light is $1/2$, respectively. Where can we find this factor $1/2$ in the expression $|r\rangle = \frac{|h\rangle + i|v\rangle}{\sqrt{2}}$? Clearly, we obtain the intensities (as usual) by calculating the squared sum of the coefficients (amplitudes), i.e.

$$\frac{1}{2} = \left(\frac{1}{\sqrt{2}}\right)^2 = \left|\frac{i}{\sqrt{2}}\right|^2 . \tag{2.19}$$

Next, we consider light whose polarization plane is rotated by ϑ. The rotation matrix, which is known to be given by $\begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}$, transforms the state $|h\rangle \cong \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ into the rotated state $|\vartheta\rangle \cong \begin{pmatrix} \cos\vartheta \\ \sin\vartheta \end{pmatrix}$.[13] Hence we have $|\vartheta\rangle = \cos\vartheta\,|h\rangle + \sin\vartheta\,|v\rangle$. The absolute square of the coefficient of $|h\rangle$ is $\cos^2\vartheta$. This is known as the *Law of Malus*[14] and gives the relative intensity of the horizontally-polarized light.

Thus we have recovered the known relationship between intensity and absolute squared amplitude: For $|A\rangle = c_1\,|h\rangle + c_2\,|v\rangle$, the (relative) intensity of e.g. $|h\rangle$ is given by $|c_1|^2$, whereby $|c_1|^2 + |c_2|^2 = 1$ has to hold. In other words: the state $|A\rangle$ must be *normalized*.

---

[13] The active rotation (rotation of the vector by ϑ counterclockwise) is given by $\begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}$; the passive rotation (rotation of the coordinate system) by $\begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix}$.

[14] Perhaps familiar from school or undergraduate laboratory courses?

2.2 Light as Photons

The above considerations are independent of the intensity of the light—they apply to an intense laser beam as well as to the dimmest glow. But if we can turn down the intensity of a light source sufficiently far, we eventually encounter a situation where the light consists of a stream of *single photons*.[15] Even then—and that is the crucial point—we assume that the above formulation remains valid. This is the above-mentioned jump from classical physics to quantum mechanics, which is not strictly derivable logically, but requires additional assumptions; there are in fact two of them. First, the existence of photons is assumed, which we take as an experimentally assured fact. Secondly, there is the assumption that expressions such as (2.17) and (2.18) retain their validity even for single photons.

Though it is not absolutely mandatory, as mentioned before, the second assumption is without any apparent alternatives—provided that light consists of a stream of photons—because we cannot draw the conclusion from the above considerations that these equations apply only above a certain number of photons. An additional degree of plausibility can be found in the fact that the wave character of light, for example in (2.17) and (2.18), never enters the arguments explicitly. And, finally, such an assumption—independently of plausibility—has to be proven by experiment, which of course has long since been done.

2.2.1 Single Photons and Polarization

We see that polarization is a property of *individual photons*. This fact is new and is by no means self-evident; thus, for individual photons we have e.g.

$$\begin{aligned} |r\rangle &= \tfrac{|h\rangle + i|v\rangle}{\sqrt{2}} \\ |l\rangle &= \tfrac{|h\rangle - i|v\rangle}{\sqrt{2}}. \end{aligned} \tag{2.20}$$

However, the interpretation *must* be *different* from the case of 'classical' light, since a photon in state $|r\rangle$ whose linear polarization is measured (i.e. with respect to $|h\rangle$ or $|v\rangle$) cannot split up into two linearly-polarized photons (how would the energy

[15] Single-photon experiments are standard technology these days. In 1952, Schrödinger declared: "We never experiment with just one electron or atom or (small) molecule. In thought-experiments we sometimes assume that we do; this invariably entails ridiculous consequences." Times have changed: Precision experiments using a single photon or a single atom are the basis of e.g. today's time standard, and modern quantum-mechanical developments such as the quantum computer rest on those 'ridiculous consequences'. We recall that photons (as far as we know) have immeasurably small dimensions and are in this sense referred to as point objects (or point particles). Although they represent light of a specific wavelength, they do not have a spatial extension on the order of the wavelength of the light.

$E = \hbar\omega$ be divided up in that case?).[16] We must assume that we can infer the *probabilities* P of finding a photon, initially in the state $|r\rangle$, in a state $|h\rangle$ or $|v\rangle$ after its passage through e.g. a linear polarizing filter, from (2.20); namely

$$P(h) = \left|1/\sqrt{2}\right|^2 = \frac{1}{2} \text{ and } P(v) = \left|i/\sqrt{2}\right|^2 = \frac{1}{2}. \tag{2.21}$$

Therefore, one must beware of interpreting the expressions (2.20) incorrectly—it is not at all the case that an $|r\rangle$ photon consists of half a horizontally- and half a vertically-polarized object. Rather, (2.20) tells us that an $|r\rangle$ photon contains two *possibilities* to present itself in a measurement as either $|h\rangle$ or $|v\rangle$—but only one of these is realized in any given measurement. Before the measurement, however, the photon is in a *superposition* of the two states. This is a very common trait of quantum mechanical systems: states can be superposed.

This superposition principle is valid for *all* states or objects described by quantum mechanics, whether we attribute to them more wave- or more particle-like character. In the macroscopic domain, the superposition of states would lead to very unusual effects—for example, in a system with the two states |cow in barn⟩ and |cow in field⟩, or in the famous example of Schrödinger's cat, namely |dead cat⟩ and |live cat⟩. Our direct daily experience does not include such superposed states, and so certain quantum-mechanical phenomena are in conflict with 'common sense' (whatever that may be, exactly). But as mentioned above, our sensory apparatus was trained by evolution under macroscopic conditions[17] and our understanding of the world is based on corresponding model concepts. No one will seriously argue that therefore, the whole of nature should operate according to these 'daily life rules' which literally permeate our flesh and blood.

So when speaking of the paradoxes of quantum mechanics, we should recognize that the real paradox is simply that the rules of quantum mechanics (which we can indeed recognize, identify and formulate) proceed according to a different pattern from our familiar daily-life rules (i.e. 'common sense'). But quantum mechanics works, and indeed it works verifiably, consistently, reproducibly, and with an amazingly high degree of accuracy—in short, according to all scientific standards, it is a successful theory. Quantum mechanics is one of the best if not *the* best-validated basic theory in physics. Of course, in spite of this the question remains as to why there are only microscopic superposition states and apparently no macroscopic ones. This is a central problem of quantum mechanics, which we will address several times in various chapters in the following.

[16] In a vacuum, photons are indivisible, and that holds also for most interactions with matter. One has to work hard to 'cut' photons. This can be achieved for example in the interaction with certain nonlinear crystals, where a single photon breaks up into two photons of lower energy (parametric fluorescence, see Appendix I, Vol. 2). Devices for polarization measurement are of course manufactured in such a way that they leave the photons unsplit.

[17] Furthermore, in 'slow' conditions—the effects of the theory of relativity are beyond our daily life experience, as well.

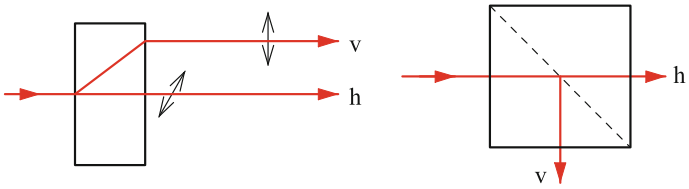

Fig. 2.3 Birefringence (*left*) and polarization beam splitter (*right*)

2.2.2 *Measuring the Polarization of Single Photons*

Back to our single photons: We produce one of them with a certain polarization and send it through the usual polarization filter (analyzer), which absorbs photons with the 'wrong' polarization,[18] or through an analyzer with two outputs, such as a birefringent crystal or a polarizing beam splitter (PBS); see Fig. 2.3. We assume an angle φ between the polarization direction and the analyzer axis. Whether a photon passes the analyzer or not, or alternatively where it leaves the PBS, can be predicted with certainty only if $\varphi = 0$ (the photon passes, or exits to the right); or $\varphi = \pi/2$ (the photon is absorbed, or exits downwards).[19] In all other cases, we can specify only the *probability* $P(\varphi)$ that the photon passes the absorbing analyzer or the PBS with the same polarization; it is given by $P(\varphi) = \cos^2 \varphi$.

We can interpret these facts as follows: Before a measurement, one cannot objectively determine whether the photon will pass through the analyzer or not. This is revealed only by the process of measurement. Which one of the two possibilities will be realized cannot be said before the measurement, but one can specify the respective probabilities of observing each of the two outcomes. We can summarize the general result: For a state such as $|z\rangle = c\,|x\rangle + d\,|y\rangle$, the term $P = |c|^2$ represents the probability of measuring the state $|x\rangle$ (assuming a normalized state $|z\rangle$, i.e. with $|c|^2 + |d|^2 = 1$). In symbolic shorthand notation,

$$\text{measuring probability} = |\text{coefficient}|^2. \tag{2.22}$$

This is a similar finding as in classical physics—with the very significant difference that the statement holds there for *intensities* , but here for *probabilities*.

We note that while it is possible to measure with sufficient accuracy the polarization which a classical wave had before the measurement, this is *in principle impossible* for a single photon of unknown polarization. Quantum objects do not always possess well-defined values of all physical quantities—a linearly-polarized photon 'has', for example, no well-defined circular polarization. If we send a

[18] We point out that this is not an exotic quantum-mechanical procedure—the eyes of every bee, or suitable sunglasses, perform precisely this kind of 'measurement process'.

[19] These cases can be produced by inserting a further analyzer whose orientation is $\varphi + 0$ or $\varphi + \pi/2$, for example.

horizontally linear-polarized photon through a linear analyzer, rotated through the angle φ, we find horizontally- and vertically-polarized light with the probabilities $\cos^2\varphi$ and $\sin^2\varphi$, respectively—in principle, with no ifs, ors, or buts.

2.2.2.1 The Ensemble

How can we verify experimentally that the calculated probabilities are correct? Obviously not in a *single* experiment. Because if we send e.g. a circularly-polarized photon through a PBS, it emerges on the other side as either horizontally or vertically polarized, and we have no information about the probabilities. So we need to repeat the measurement more than once. Now the term *ensemble* comes into play. In quantum mechanics, this term refers to a set of (strictly speaking) infinitely many identically-prepared copies of a system.[20] It is a fictitious set which has no counterpart in physical reality, but serves only for conceptual clarification. The 'strictly speaking' in brackets refers to the fact that often (and for practical reasons), N identical copies of an system are called an ensemble, if N is sufficiently large (even though not infinite).

The concept 'ensemble' allows us to calculate the probabilities for the occurrence of certain measurement values, and thus to predict them—they are given simply by the fraction of subsets of the ensemble which are characterized by the presence of those values. In the example given above, this is $P(\varphi) = \cos^2\varphi$ for passing the analyzer or $\sin^2\varphi$ for not passing through it.

We emphasize that the use of the word *ensemble* does *not* imply that the relevant physical quantities (here the polarization) have well-defined values which are distributed in some unknown way among the members of the ensemble. In the example, the ensemble consists of horizontally-polarized photons whose polarization properties are *not* defined with respect to an analyzer that is rotated by φ.

In practice, one can of course measure only finitely many systems; often one has to make do even with a *single* system (in the example, a single photon). But the predictions arising from the concept of ensembles are generally valid and apply (in terms of probabilities) to the particular case considered.[21]

So we can imagine that we prepare N systems in an identical manner and always measure the same variable, in our example the rate of observation of vertically- or horizontally-polarized photons behind the PBS.[22] For $N \to \infty$, the relative frequencies of occurrence of the different measurement results become the probabilities of the ensemble; in the above example $\cos^2\varphi$ and $\sin^2\varphi$.

---

[20] The systems need not be in the same state, but the preparation process must be the same.

[21] Just as the interference pattern in the double slit experiment builds up gradually from scattered spots over time.

[22] Another example of an ensemble are electrons which are prepared by a Stern–Gerlach apparatus and a velocity filter so that their spins are pointing upwards and their speeds are confined to a particular interval $(v - \Delta v, v + \Delta v)$. A further example is a set of hydrogen atoms in a particular excited state, whereby here the preparation refers to the energy, but not to the angular momentum of the state.

2.2.2.2 Ensemble or Single Object?

As we can see, the experimental verifiability of the theory is not guaranteed for a single quantum object, but rather requires an ensemble. Hence, one can argue that the formalism developed above (regardless of our derivation) is essentially a mathematical rule which applies only to an ensemble (so-called ensemble interpretation). Another position asserts however that the formalism applies also to an individual quantum object, as we have assumed to be the case. Both interpretations lead to the same results, but they are based on different concepts of 'reality'.

We encounter here for the first time a situation typical of quantum mechanics: The formalism and the verification of its predictions by measurements are uncontroversial (if we accept certain basic assumptions); the controversial issue concerns what quantum mechanics 'really' means. This debate is as old as quantum mechanics itself, and is still very much alive; there are a dozen or more different explanations (interpretations). We will discuss these questions often and will give an overview of current interpretations in Chap. 28, Vol. 2.

2.2.2.3 Do We Really Need Probabilities?

Finally, a remark about the concept of probability. In classical physics, probabilities reflect the fact that we do not know (or do not wish to know) enough about some of the properties of a system in order to calculate them explicitly. For instance, in the kinetic theory of gases, one is not interested in the behavior of a single molecule; a well-known example from a quite different field are the opinion polls before an election, where the behavior of individual voters is not of interest. Analogously, one could assume here that the occurrence of probabilities indicates that below the level of our discussion, there are some *hidden variables*, and if we were able to know them, we could formulate the whole process exactly without resorting to probabilities. This is an obvious idea which was brought up very soon after the emergence of quantum mechanics. It took nearly 40 years until a criterion was found to resolve this question in principle, and a few more years to disprove the idea of hidden variables experimentally based on that criterion—at least this holds for the major classes of hidden variables. More on this topic in Chaps. 20 and 27, Vol. 2.

According to our current knowledge, we cannot avoid the term 'probability' in quantum mechanics. It is, so to speak, a structural element of quantum mechanics, the sign that quantum mechanics deals first of all with possibilities, one of which is realized by a measurement, with a certain probability for different outcomes.

2.3 Exercises

1. Given an electromagnetic wave $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{i(\mathbf{kr}-\omega t)}$ in a charge-free region of space (we consider the electric field only); show that the wave is transverse, i.e. that $\mathbf{k} \cdot \mathbf{E}_0 = 0$ holds (Hint: cf. the Maxwell equation $\nabla \mathbf{E} = 0$). Specialize to $\mathbf{k} = (0, 0, k)$.
2. Linear combinations
   (a) Express $|r\rangle$ as a linear combination of $|h\rangle$ and $|v\rangle$. Do the same for $|l\rangle$.
   (b) Express $|h\rangle$ as a linear combination of $|r\rangle$ and $|l\rangle$. Do the same for $|v\rangle$.
3. A phase shift of 90° is described by $e^{i\pi/2} = i$. What follows for a phase shift of 180°?
4. Elliptical polarization: Given the state $|z\rangle = \alpha\,|h\rangle + \beta\,|v\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$; express $|z\rangle$ as a superposition of $|r\rangle$ and $|l\rangle$.